TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

OUTI HAAPANEN

PROTON TRANSLOCATION CHANNELS IN RESPIRATORY COMPLEX I PROBED BY MOLECULAR DYNAMICS SIMULATIONS

Master of Science thesis

Examiner: Prof. Ilpo Vattulainen
Examiner: Dr. Sc. Simo Ali-Löytty
Supervisor: Vivek Sharma, PhD
Examiners and topic approved by the Faculty Council of the Faculty of Natural Sciences
on 4th May 2016

ABSTRACT

OUTI HAAPANEN: Proton Translocation Channels in Respiratory Complex I Probed by Molecular Dynamics Simulations
Tampere University of Technology
Master of Science thesis, 72 pages, 10 Supplementary pages
December 2016
Master's Degree Programme in Science and Engineering
Major: Mathematics
Examiners: Prof. Ilpo Vattulainen and Dr. Sc. Simo Ali-Löytty
Supervisor: Vivek Sharma, PhD
Keywords: respiratory chain, complex I, proton pumping, water dynamics, molecular dynamics simulations

Energy production is one of the vital functions in living cells. In oxidative phosphorylation, the nutrients in foodstuff are converted to the energy currency of cells, adenosine triphosphate (ATP). Five respiratory complexes embedded in the inner mitochondrial membrane in eukaryotes, and in the inner cell membrane of bacteria, perform oxidative phosphorylation. The first enzyme in the respiratory chain, NADH(nicotinamide adenine dinucleotide)-ubiquinone oxidoreductase (complex I) is one of the largest known protein assemblies. Complex I is L-shaped and consists of two domains: the hydrophilic domain in the matrix and the hydrophobic domain in the membrane. The enzyme transfers two electrons from NADH to quinone, utilises the released free energy in pumping four protons across the membrane and thus participates in the generation of the proton electrochemical gradient. Dysfunctions and mutations in complex I have been connected to several incurable neurodegenerative diseases, for instance Alzheimer's disease. Understanding the molecular function of complex I is the key to unravel the dysfunctions and develop cures against these lethal diseases. The mechanism of complex I is elusive, and many questions concerning the coupling mechanism and proton pumping remain unsolved. The objective of this thesis is to study the possible proton translocation pathways in the antiporter-like subunits Nqo12-14 in the membrane arm of complex I. The research question is addressed using a computational method, known as atomistic molecular dynamics (MD) simulations. The results reveal a set of conserved hydrophilic residues coordinating the water wire formation. Additionally, a strong water mediated connection through the middle plane of the membrane arm in subunits Nqo12-14 is observed, giving rise to a tentative coupling element. Water dynamics along the long horizontal helix HL unique in subunit Nqo12 is also analysed, and a stabilising mechanism of subunits Nqo12-14 is proposed.

TIIVISTELMÄ

OUTI HAAPANEN: Protonien kuljetuskanavat hengitysketjun kompleksissa I molekyylidynamiikkasimulaatioilla tutkittuna
Tampereen teknillinen yliopisto
Diplomityö, 72 sivua, 10 liitesivua
Joulukuu 2016
Teknis-luonnontieteellinen koulutusohjelma
Pääaine: Matematiikka
Tarkastajat: Prof. Ilpo Vattulainen ja TkT Simo Ali-Löytty
Ohjaaja: Vivek Sharma, FT
Avainsanat: hengitysketju, kompleksi I, protonipumppaus, vesidynamiikka, molekyylidynamiikkasimulaatiot

Energiantuotanto on yksi solujen elinehdoista. Ravintoaineiden energia muutetaan solujen energianlähteeksi, adenosiinitrifosfaatiksi (ATP), oksidatiivisen fosforylaation avulla. Eukariooteissa oksidatiivinen fosforylaatio tapahtuu viiden entsyymikompleksin muodostamassa hengitysketjussa mitokondrion sisäkalvolla. Prokariooteissa vastaavat kompleksit sijaitsevat sisemmällä solukalvolla, koska erillistä mitokondriota ei niissä ole. Hengitysketjun ensimmäinen entsyymi, NADH(nikotiiniamidiadeniinidinukleotidi)-ubikinonioksidoreduktaasi eli kompleksi I, on eräs suurimmista tunnetuista proteiineista. Se on L-kirjaimen muotoinen ja koostuu kahdesta osasta: hydrofiilinen osa, joka sijaitsee mitokondrion matriisissa, ja hydrofobinen osa, joka sijaitsee kalvossa. Kompleksi I siirtää kaksi elektronia NADH:lta kinonille, ja tässä vapautuva vapaaenergia käytetään neljän protonin pumppaamiseen kalvon läpi. Täten kompleksi I on osallisena protonien muodostaman elektrokemiallisen gradientin luomisessa. Toimintahäiriöt ja mutaatiot kompleksi I:ssä on yhdistetty useisiin parantumattomiin neurodegeneratiiviisiin sairauksiin kuten Alzheimerin tautiin. Kompleksi I:n normaalin toiminnan ymmärtäminen on avain toimintahäiriöiden ymmärtämiselle ja tulee mahdollistamaan uusien hoitomuotojen kehittämisen näitä tällä hetkellä tappavia sairauksia vastaan. Kompleksi I:n toimintamekanismi ei ole täysin selvillä, mikä tuo esille useita kysymyksiä koskien muun muassa protonipumppausta. Tämän diplomityön tavoitteena on tutkia protonipumppauskanavia kompleksi I:n hydrofobisen osan yksiköissä Nqo12-14. Tutkimuskysymystä selvitetään laskennallisesti hyödyntäen atomaarisia molekyylidynamiikkasimulaatioita. Tutkimuksessa löydetään joukko hydrofiilisiä aminohappoja, jotka osallistuvat vesikanavien muodostamiseen ja siten protonipumppaukseen. Lisäksi simulaatioista paljastuu voimakas vesimolekyyliyhteys tutkittavien yksiköiden välillä kalvon keskitasossa valottaen mahdollista kinonin pelkistämisen ja protonien pumppauksen yhdistämismekanismia. Yksikkö Nqo12 sisältää uniikin rakenteen, jonka analyysi paljastaa myös mahdollisen stabilointimekanismin kompleksi I:n kalvo-osalle.

PREFACE

This thesis was written during the summer 2016 in the Department of Physics at the Tampere University of Technology. The research presented here was conducted between fall 2015 and summer 2016. The objective of this thesis was to study the water dynamics in the antiporter-like subunits Nqo12-14 in the membrane domain of respiratory complex I. The research question was addressed computationally using molecular dynamics simulations. The computing resources for this work were provided by CSC (Center for Scientific Computing) - IT center for science.

First of all, I want to thank my supervisor Doctor of Philosophy Vivek Sharma, who has patiently helped me and answered my questions for all the years we have worked together. His help and support in the research and in the writing process have been most valuable, and I very much look forward to our future collaborations. I also wish to thank the Biological physics and Soft matter group leader Professor Ilpo Vattulainen for giving me the opportunity to join his team, to work on interesting topics, and to challenge myself. Additionally, I would like to thank the Department of Mathematics at the Tampere University of Technology and Doctor of Science Simo Ali-Löytty for accepting and understanding my peculiar choice of thesis topic.

Last but not least, I express my profound gratitude to all the people that have supported me through my studies and the completion of this thesis: family, friends, and co-workers.

Tampere, December 2016

Outi Haapanen

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

2D	two-dimensional
3D	three-dimensional
ADP	Adenosine diphosphate
APL	Area per lipid
ATP	Adenosine triphosphate
CG	Conjugate gradient
DNA	Deoxyribonucleic acid
ETC	Electron transport chain
FMN	Flavin mononucleotide
MD	Molecular dynamics
MSA	Multiple sequence alignment
NADH	Nicotinamide adenine dinucleotide
NAMD	Nanoscale molecular dynamics
OXPHOS	Oxidative phosphorylation
P_i	Inorganic phosphate
PBC	Periodic boundary conditions
PDB	Protein data bank
PME	Particle mesh Ewald
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
Q	Quinone
RNA	Ribonucleic acid
mRNA	Messenger RNA
tRNA	Transfer RNA
ROS	Reactive oxygen species
STAMP	Structural alignment of multiple proteins
VMD	Visual molecular dynamics

$\mathbf{a}(t)$	acceleration
$\mathbf{b}(t)$	first derivative of acceleration
$\mathbf{c}(t)$	second derivative of acceleration
$f(x)$	potential energy function
$\mathbf{F}_i$	force on particle i
$\mathbf{g}_k$	gradient

i	index
j	index
k	step in conjugate gradient algorithm
$k_{f,ij}$	spring constant for harmonic bond stretching potential
$k_{f,\theta}$	force constant for harmonic angle bending potential
$k_{f,\phi}$	force constant for dihedral angle torsion potential
M	mass matrix
n_i	multiplicity of dihedral torsion angle
N	normalization factor in Q_{res} computation
N	number of atoms in RMSD calculation
N_{res}	number of residues in protein
N_{seq}	number of proteins
$\mathbf{p}_k$	search vector in conjugate gradient algorithm
q	charge
Q	native contacts in protein
Q_{res}	degree of structural similarity
r_{ij}	distance between two atoms
$r_{f,ij}$	reference distance between two atoms
$r_{ij}^{(n)}$	distance between Cα atoms between residues i and j in protein n
$\mathbf{r}(t)$	position
S_{CD}	deuterium order parameter
Δt	timestep of integration
U_{angle}	potential energy of angle bending
U_{bond}	potential energy of covalent bond stretching
U_{bonded}	potential energy of bonded interactions
$U_{Coulomb}$	potential energy of electrostatic interactions
$U_{dihedral}$	potential energy of torsion angle bending
$U_{non-bonded}$	potential energy of non-bonded interactions
U_{total}	total potential energy
U_{vdW}	potential energy of van der Waals interactions
$\mathbf{v}(t)$	velocity
x_i	position of particle i
x_0	initial configuration
x_k	configuration at step k
x_{k+1}	configuration at step $k+1$
β_{k+1}	conjugate gradient parameter

ϵ_0	electric constant
ϵ_{ij}	depth of potential well
θ_i	angle between three atoms
$\theta_{f,i}$	reference angle between three atoms
λ_k	step length
$\Delta\tilde{\mu}_{H^+}$	proton electrochemical gradient
σ_{ij}	distance at which the potential between two atoms is zero
σ_{ij}^2	variance term in Q_{res} value computation
ϕ_i	torsion angle between four atoms
$\phi_{f,i}$	reference torsion angle between four atoms
$\Delta\Psi$	electrical potential difference
ω	angle between carbon-hydrogen bond and membrane normal

1. INTRODUCTION

Sun is the ultimate source of energy for all living matter. Photosynthesis is performed by plants and algae to produce glucose and oxygen from sunlight and carbon dioxide. Living matter consumes glucose in food as such, or as bigger, more complex compounds. The organisms collect and store the nutrients from food, which are later released via metabolism. There are several forms of energy conversion processes depending on the available nutrients and the environment. The goal of different processes is to produce electron-donating substrates, such as pyruvate in glycolysis [1, p. 492]. Adenosine triphosphate (ATP) is the energy currency of the cells and crucial to cell function. ATP is efficiently produced via oxidative phosphorylation (OXPHOS) [1, p. 563], which requires a proton electrochemical gradient generated by the electron transport chain (ETC). The production of ATP via ETC takes place in mitochondria and bacteria. Another process to produce ATP is glycolysis in mitochondrial matrix or bacterial cytosol, but it is far less efficient than OXPHOS. In OXPHOS glucose is completely oxidised, hence OXPHOS produces approximately 30 ATP molecules per glucose molecule, whereas glycolysis only produces 2 ATP molecules per glucose molecule [2, p. 464].

NADH-ubiquinone oxidoreductase (respiratory complex I) is the first and the largest enzyme in the electron transport chain. The function of complex I is to transfer two electrons from NADH to quinone and to couple the released free energy to pump four protons across the inner mitochondrial membrane. The structure of the hydrophilic domain of complex I was published 10 years ago [3], and the structure of the entire protein was published only recently in 2013 [4]. Crystal structures revealed a unique property of this fascinating molecular machine: the membrane arm contains no redox-centers. This observation rewrote theories regarding the mechanism of complex I, as without redox-centers a specific coupling mechanism between quinone reduction and proton pumping must exist [5, p. 111]. Two types of coupling mechanisms have been proposed: redox-driven (referred as direct), and conformation-driven (referred as indirect).

Complex I is one of the major reactive oxygen species (ROS) producers. Additionally, it has been suggested to play a role in several neurogenerative diseases, such as Parkinson's disease and Alzheimer's disease, as well as in ageing. Understanding how complex I functions will bring up new possibilities in finding cures for these diseases.

This work elucidates the proton pumping and water dynamics in the membrane arm of complex I. Two independent 500 ns long atomistic molecular dynamics (MD) simulations on the crystal structure of the entire bacterial complex I are performed, and the water dynamics in subunits Nqo12-14 are analysed in detail. Additionally, sequence and structural analysis is performed on the enzyme from different organisms to reveal the conserved residues as well as to provide means to compare the results to earlier studies, which were based on the crystal structures of the enzyme from different species.

This thesis is structured as follows. Chapter 2 introduces the biological background of this study starting from the basics of cell biology. Chapter 3 reviews the theory of the used computational technology and software. Simulation details, model systems, and analysis tools are presented in Chapter 4. Chapter 5 presents and discusses the results, and Chapter 6 summarises the key findings and their significance.

2. BIOLOGICAL BACKGROUND

This Chapter introduces the biological background of this thesis work. The overview begins with the basic components of living matter: atoms, biomolecules and cells. Mitochondria is introduced prior to the description of the respiratory chain and the detailed portrait of respiratory complex I - the enzyme studied in this work.

2.1 Building blocks of living organisms

Living matter is mainly composed of six elements: carbon, hydrogen, oxygen, nitrogen, phosphorus and sulphur [6, p. 46]. Even though these elements constitute the majority of atoms in all organisms, other elements, for instance sodium, are present in smaller amounts and often in ionic form. Different combinations of atoms can form various chemical compounds, some of which are essential to life, while some may even be toxic. Compounds essential for all living organisms can roughly be categorized into four main classes: nucleotides, amino acids, fatty acids and sugars [6, p. 49]. Water is the biological solvent, and in an aqueous environment phosphoric acid dissociates to inorganic phosphate (HPO_4^{2-}), called P_i, and two protons [6, p. 46]. Compounds can be combined in multiple ways to constitute large biomolecules, such as deoxyribonucleic acid (DNA), proteins and lipids. Biomolecules form intricate structures, such as cell organelles and membranes, which constitute the unit of life, the cell.

Genetic information is stored in DNA. DNA is a large double-helical molecule, which is composed of nucleotides. Nucleotides are composed of phosphate, deoxyribose and base. The nucleotides form specific pairs based on the bases: cytosine - guanine and adenine - thymine pairs. DNA includes the information required to produce all proteins vital to cell function. The first step in protein manufacturing is transcription, in which the information in the gene (DNA) is copied into single-stranded form, messenger ribonucleic acid or mRNA. mRNA is also composed of nucleotides, but

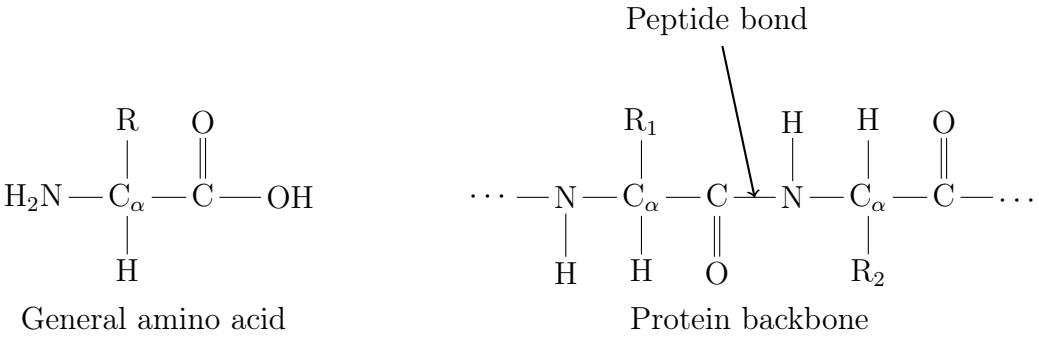


Figure 2.1 *Amino acids and protein backbone. Peptide bond forms between the carboxyl and the amino group of two amino acids. R stands for side chain.*

instead of deoxyribose, mRNA includes ribose. The bases of RNA can form pairs: cytosine - guanine and adenine-uracil pairs. However, as mRNA is single stranded, no pairs are spontaneously formed. mRNA is turned into protein in a process called translation. The process functions with a key-lock principle, in which codons are keys and mRNA the lock. Codons are groups of three bases, which correspond to amino acids. Transfer RNA, tRNA, are special carriers, which include a codon and amino acid. Ribosomes are large molecular machines, which trap tRNA from cytosol. Ribosome moves along the mRNA and holds the captured tRNA molecule in position. If the codon pairs with the bases in mRNA, tRNA releases its amino acid, and the ribosome covalently bonds it to the amino acid chain being created. The formed amino acid chain is then transferred for further processing into other cell organelles, such as the Golgi apparatus, to make a functional protein.

Proteins are composed of amino acids. All amino acids have a similar structure: Cα atom, amine group NH_2, carboxyl group COOH and the side chain core [7, p. 4-5]. Figure 2.1 shows the basic structure of an amino acid. There are 20 different amino acids, and the side chain is what distinguishes amino acids from each other. Based on the properties of the side chain, amino acids can be categorized into three classes:

- *hydrophobic*: alanine, valine, leucine, isoleucine, phenylalanine, proline, methionine
- *charged*: aspartic acid, glutamic acid, lysine, arginine
- *polar*: serine, threonine, cysteine, asparagine, glutamine, histidine, tyrosine, tryptophan

The 20th amino acid, glycine, is special as it has two hydrogen atoms attached to its Cα atom, hence no sidechain. Amino acids joined by peptide bonds form amino acid sequences, shown in Figure 2.1. Knowing the sequence of a protein, that is the primary structure, is however not enough in understanding the structure and function, as amino acids can be arranged in numerous ways (protein structures) in space. For example, if an amino acid sequence contains 100 amino acids, and each amino acid has two possible arrangements, the number of possible conformations is roughly 2^{100}. The secondary structure of a protein includes alpha-helices and beta-sheets. The tertiary structure shows how the polypeptide is folded into a three-dimensional (3D) structure. Finally, the quaternary structure displays the arrangement of multiple domains, except for very small sequences. The 3D structures cannot be modelled accurately from protein sequences, rather they are resolved using for example X-ray crystallography [7, p. 4]. Secondary structures may serve functional purpose. For example, in respiratory complex I a long horizontal helix HL in subunit Nqo12 and beta-sheets βH in subunits Nqo12-14 are conserved, and have been proposed to play a role in stabilising the membrane domain [8]. See Figure 2.5b for the location of these structures. Secondary structures are characterized by two torsion angles in the backbone: ϕ is the torsion angle between C-N-Cα-C, and ψ the torsion angle between N-Cα-C-N. See Figure 2.1 for arrangement of the backbone and these angles. To study the secondary structure, Ramachandran plots are often used. Ramachandran plot shows the ϕ and ψ angles of amino acids in the protein. Figure 2.2 shows the Ramachandran plot of subunit Nqo12 in the crystal structure of complex I obtained from *Thermus thermophilus*.

Lipid molecules are made of fatty acids. They have a hydrophilic head and hydrophobic tail [9, p. 336], hence they are amphiphilic in nature [6, pp. 315-316]. A characteristic of amphiphiles is their tendency to self-assemble in water. The lipids self-assemble in such a way that the hydrophilic heads are in contact with water, while the hydrophobic tails are in the centre [9, p. 336]. The shape of the assembled structure depends on the geometry of the lipid. If the lipids are conical, then micelles or vesicles are formed, but if the lipids are cylindrical, then they prefer to assemble in a bilayer [10, pp. 81-82]. Phospholipids are a biologically relevant class of lipids, because they are present in most animal cell membranes [10, p. 290]. All phospholipids share similar features. Glycerol is the centre to which two fatty acid chains and a phosphate are covalently bonded [10, p. 290]. Very often a hydrophilic molecule, for instance choline, is covalently bonded to the phosphate. Additionally,

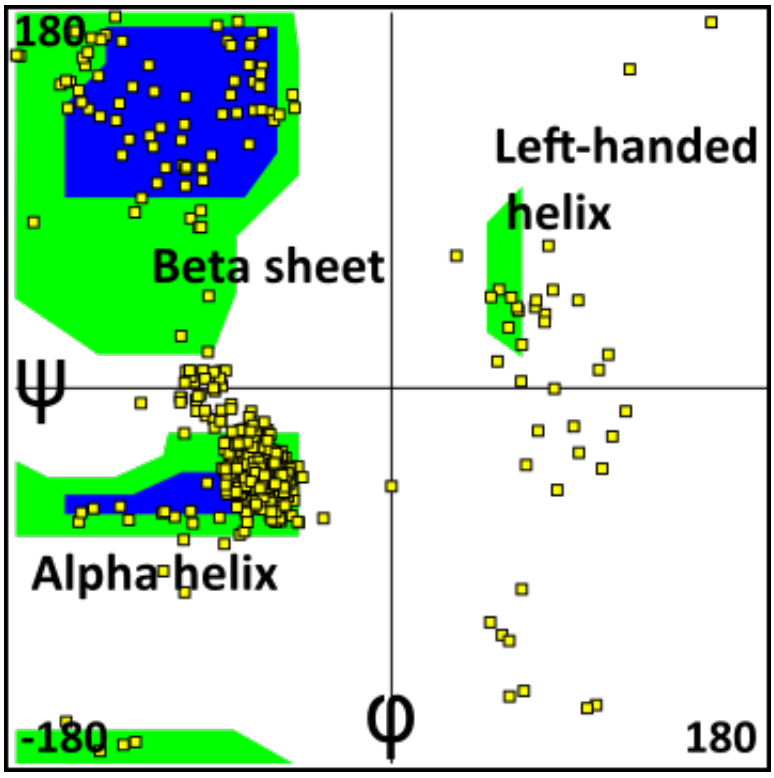


Figure 2.2 *Ramachandran plot of subunit Nqo12. Most of the residues form alpha helical structures.*

several phospholipids have one saturated and one unsaturated acyl chain, which decreases the chain melting temperature below the physiological temperature. Figure 4.1 shows the structure of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid. Phospholipid bilayers are stable, fluidic and flexible, which makes them ideal for compartmentalization [9, pp. 338-339].

Cells are the structural and functional basic units of living organisms [2, p. 1]. They carry out a large number of functions, for example synthesizing new molecules, storing genetic information, reacting to changes in the environment, removing waste products and converting energy to other usable forms. Cells are small in size, and small entities have a higher surface area to volume ratio, which enables for example efficient exchange of nutrients [11, p. 69]. The size of cells is typically within the μm scale [6, p. 40]. Cells can be classified to prokaryotic, which do not contain a nucleus, and eukaryotic, which do [6, p. 40]. Prokaryotes are larger than eukaryotes. Bacteria and archaea constitute the prokaryotes and plants, animals and fungi the eukaryotes. Prokaryotic cells are simpler in comparison to eukaryotic cells. In addition to the lack of nucleus, prokaryotes do not include all the cell organelles

eukaryotes do, for example the mitochondria [6, p. 40] .

Mitochondria are cell organelles often referred to as the powerhouses of cells. It is hypothesized that they have evolved from alpha-proteobacteria [12], hence the resemblance between the mitochondria and bacteria. Similar to their bacterial ancestor, mitochondria have a double membrane structure, with inner and outer membranes separated by the intermembrane space. The inner and outer membranes differ vastly in composition and structure [2, pp. 456-457]. The outer membrane is porous due to a high number of porins, which are transport proteins [13]. They create water-filled channels through the membrane allowing transportation of molecules up to 5000 Daltons [2, p. 457] [14]. As the outer membrane is porous and permeable, membrane potential across it cannot be generated [13]. On the contrary, a membrane potential of approximately 180 mV is present across the inner membrane [13]. The inner membrane is structurally different from the outer membrane, because it forms a tight barrier impermeable to all ions and molecules [13]. Transportation across the inner membrane takes place via selective transport proteins, which effectively regulates the environment inside the mitochondria [2, p. 457]. Additionally, the enzymes involved in energy production are embedded in the inner membrane. In order to effectively increase the surface area of the inner membrane, it forms structures, which fold into the mitochondrial matrix, called cristae. In all mitochondria, cristae take up the majority of the inner membrane, however the number and packaging of cristae depend on the cell type [13]. In general, the higher the energy demand, the denser cristae are packed. This increases the surface area of the inner membrane significantly, thus allowing a higher number of respiratory enzymes. For example, in heart tissue, which has a high energy demand, cristae are closely packed [13].

The double-membrane structure of mitochondria creates two separate compartments: mitochondrial matrix and intermembrane space. Due to the differences in membrane structure, the compartments have different chemical composition and conditions. The inner membrane envelops mitochondrial matrix, and both membranes confine the intermembrane space. As the inner membrane is impermeable, the constitution of the matrix is tightly controlled. The pH of the matrix is in the range of 7.9 to 8.0 [15], whereas the pH of the intermembrane space is in the range of 7.2-7.4 [13]. The mitochondrial matrix includes the genetic material of mitochondria and the machinery required to process and duplicate it as well as enzymes involved in glycolysis and the citric acid cycle. Therefore, the protein density of the

matrix is high, up to 270-560 mg protein/ml [16, 17]. Mitochondria have their own DNA, which makes mitochondria unique among cell organelles. Mitochondrial DNA, mtDNA, is circular and composed of ca. 16,500 base pairs coding 13 mitochondrial proteins [18]. The number of different proteins in eukaryotic mitochondria is however over 1,500 [19–22], therefore mitochondria work in cooperation with nucleic DNA. However, mitochondria divide independently of their host cell, thus upon cell division the daughter cells receive intact mitochondria from the parent. In addition to energy production, mitochondria have several other functions as well. Mitochondria are involved in several processes connected to ATP synthesis in eukaryotic cells, for example autophagy, calcium signalling and apoptosis [18, 23]. Nevertheless, mitochondria are best known for energy production via the respiratory chain, which converts the chemical energy of foodstuff into molecules of ATP.

2.2 The mitochondrial respiratory chain

The mitochondrial respiratory chain consists of respiratory complexes I-V, which are embedded in the inner mitochondrial membrane of eukaryotic cells. In bacteria, the respiratory chain is located in the bacterial inner cell membrane, as prokaryotes lack the mitochondria. The respiratory chain can be considered to comprise two types of "pumps": the primary and the secondary. A primary pump is one, which creates the proton electrochemical gradient across the inner mitochondrial membrane as shown in Figure 2.3. The negative side, called the N-side, is in the mitochondrial matrix, whereas the positive side, called the P-side, is in the intermembrane space. The secondary pump is complex V, ATP synthase, which synthesizes ATP from ADP and P_i. The proton electrochemical gradient is vital for ATP production, as without it complex V would pump protons across the membrane in the same direction as the ETC, and no ATP would be generated [5, p. 4]. However, the proton electrochemical gradient is also utilised in other processes, such as mitochondrial protein import [24]. The proton electrochemical gradient, marked $\Delta\tilde{\mu}_{H^+}$, consist of two components: the concentration difference ΔpH and the electrical potential difference $\Delta\Psi$ [5, p. 4]. In mitochondria, $\Delta\Psi$ is the primary component, because this allows the enzymes on both sides of the inner membrane function at a pH that is close to neutral. The gradient is generated via proton pumping.

Three of the complexes in ETC are proton pumps: complex I, complex III, and complex IV. Complex II is involved in the Krebs cycle and does not pump protons

across the membrane.

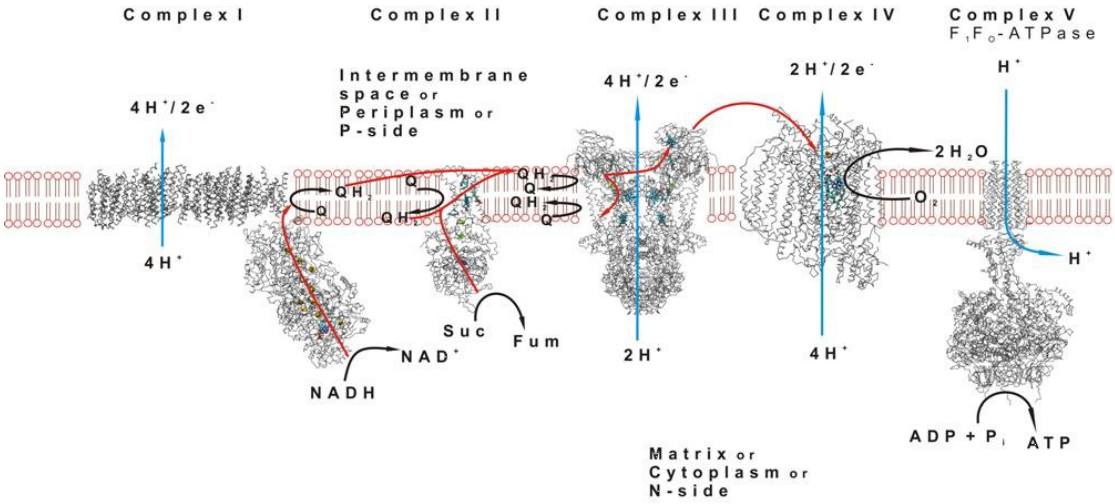


Figure 2.3 *The respiratory chain. With permission from reference [25].*

NADH-ubiquinone oxidoreductase, also called respiratory complex I, is the first enzyme in the respiratory chain. Complex I is a reversible proton pumping enzyme, and it couples the reduction of ubiquinone to proton pumping, with $4H^+/2e^-$ stoichiometry. Complex I is discussed in detail in Section 2.3.

Complex II (succinate dehydrogenase) is not a proton pump but a four-subunit-enzyme connected to the Krebs cycle [5, p. 116]. It turns succinate into fumarate and at the same time releases two protons to the N-side. Succinate releases two electrons to FAD, which are then transferred to a haem via iron-sulphur clusters. The two electrons in the haem are used to reduce quinone to quinol; however, the exact mechanism is elusive. In fact, there are two known versions of complex II: with one haem and with two haems [5, p. 116].

Complex III (ubiquinol-cytochrome *c* oxidoreductase) is the second proton pump in the mitochondrial respiratory chain. Complex III transfers electrons from ubiquinol to cytochrome *c*, and pumps protons across the membrane [5, pp. 118-123]. The thermodynamic ratio for proton pumping in complex III is 1 H^+/e^-, but quinol oxidation close to the P-side of the membrane releases one extra proton. Similar to the other respiratory proteins, complex III includes iron-sulphur clusters and two different types of haems. A functional form of complex III is a homodimer, and each monomer consists of as many as 11 subunits. However, only three of these

form the catalytic core, one may have a role in assembly and the rest are accessory. Therefore, bacterial complex III has less subunits than the mitochondrial one, like in the case of complex I (see below).

Complex IV (cytochrome *c* oxidase) is the third and the final proton pump in the electron transport chain. It belongs to the haem-copper oxidase super family [26]. Complex IV reduces oxygen to water and couples this to the translocation of protons across the membrane with a thermodynamic efficiency of $2H^+/e^-$ [5, pp. 126-131].

Altogether, according to the current consensus values, complexes I, III and IV pump 10 protons across the membrane per $2e^-$. This creates the proton electrochemical gradient across the membrane. F_1F_o-ATP synthase utilises the electrochemical gradient and produces ATP from ADP and one phosphate P_i. ATP-synthase is the final step in the mitochondrial respiratory chain as it performs the actual production of ATP. It has been proposed that in the inner mitochondrial membrane, ATP synthase molecules are arranged as dimers that act as building blocks of larger respiratory assemblies in the cristae of the membrane [27]. Adenosine triphosphate, ATP, is an "energy-rich" compound composed of three molecules: adenine, ribose and three phosphates. The name adenosine refers to adenine and ribose together [2, p. 57]. The structure of ATP is shown in Figure 2.4. The three phosphates are connected by two phosphoanhydride bonds, which can be broken by hydrolysis. The breakage of the bond releases energy, which can be utilised in cellular processes [2, p. 57]. The molecule with adenine, ribose and two phosphate groups is called adenosine diphosphate, ADP. The energy cycle of ATP and ADP is shown in Figure 2.4. ATP is used as energy currency in nearly all energy-requiring biochemical reactions, thus cells maintain a high concentration of ATP at all times [6, p. 46]. However, ATP cannot be stored; rather it has to be consumed immediately after production. Therefore, the cellular energy production must be continuous and effective.

Even though the respiratory chain is often presented as a chain-like structure (see Figure 2.3), it has been proposed that the respiratory complexes in fact assemble into units called supercomplexes or respirasomes, which have been biochemically isolated, characterized and reviewed in several studies [28–31]. Assembly to even higher-ordered structures has been proposed [32]. The role and meaning of supercomplex organisation has been studied, but remains controversial [33]. It has been

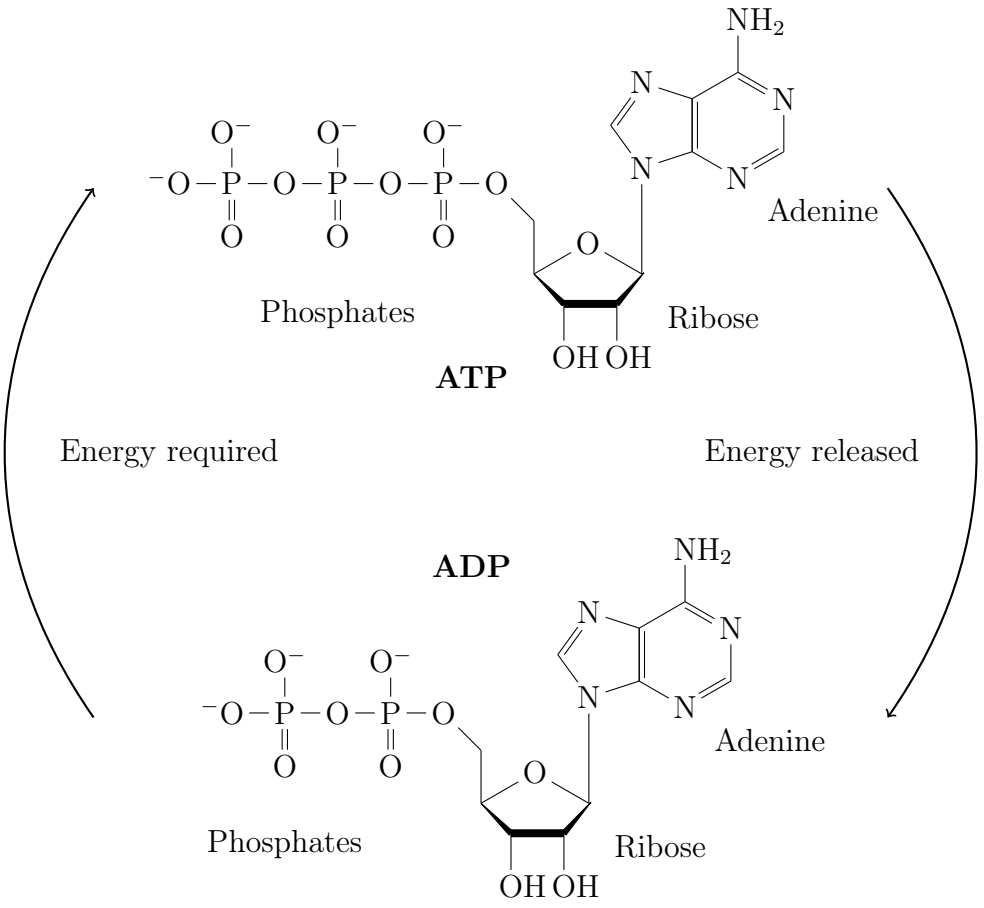


Figure 2.4 *ATP/ADP energy cycle. ATP is composed of adenine, ribose and three phosphates. ADP has the same components, but only two phosphates. The phosphoanhydride bond between the last and second last phosphates in ATP is a "high-energy bond". Breaking this bond releases energy, which can be utilised in cellular function. ATP synthase produces ATP from ADP utilising the energy of the proton electrochemical gradient.*

suggested that organisation to supercomplexes may provide kinetic advantages in electron transfer, and protect cells from oxidative stress by limiting the reactive oxygen species production of complex I [33].

2.3 Complex I

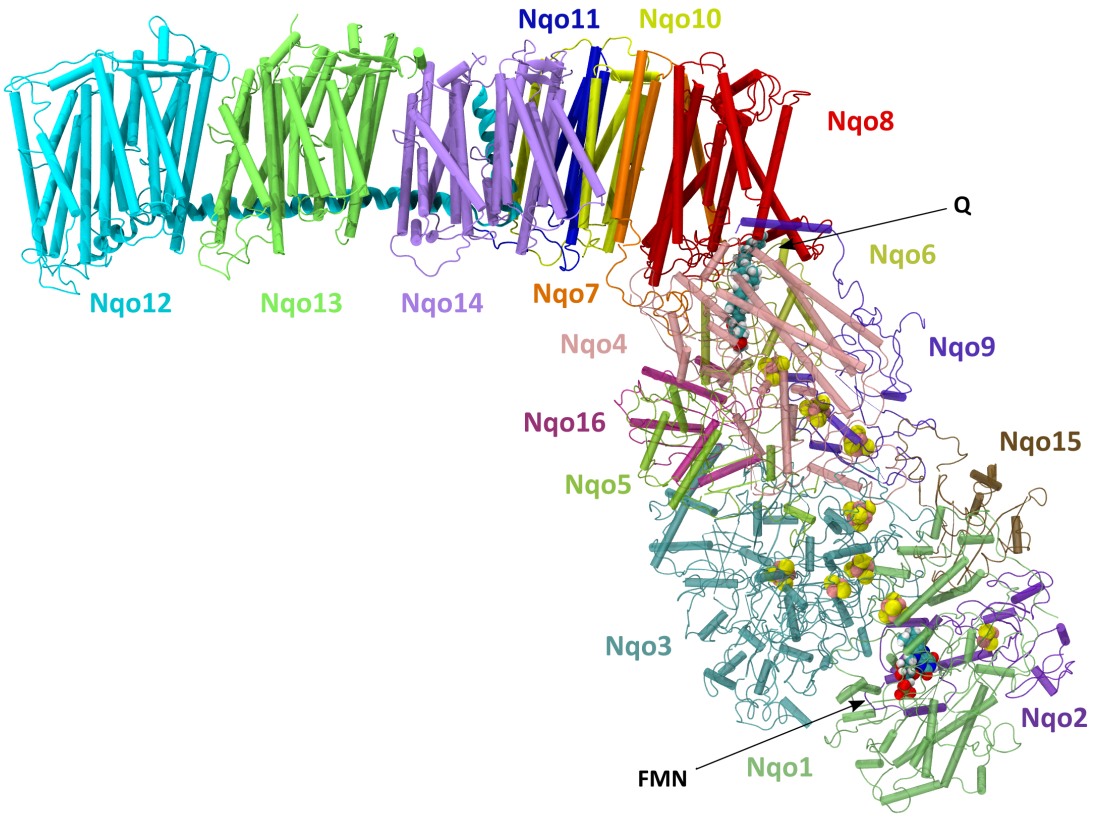
NADH-ubiquinone oxidoreductase, also called respiratory complex I, is the first enzyme in the respiratory chain. The molecular weight of complex I varies from 550 kDa to 1 MDa depending on the organism, which makes complex I one of the biggest known protein assemblies [34, p. 4]. The crystal structure for complex I remained

elusive until about 10 years ago, when the first crystal structure of the hydrophilic domain was resolved (PDB id 2FUG) [3]. Since then remarkable work has been done in determining the crystal structure of the entire complex I both from bacteria as well as eukaryotes [3,4,8,35–38]. The crystal structure used in this work is bacterial complex I from *Thermus thermophilus* (PDB id 4HEA) [4]. Figure 2.5 shows the structure of complex I from *Thermus thermophilus*.

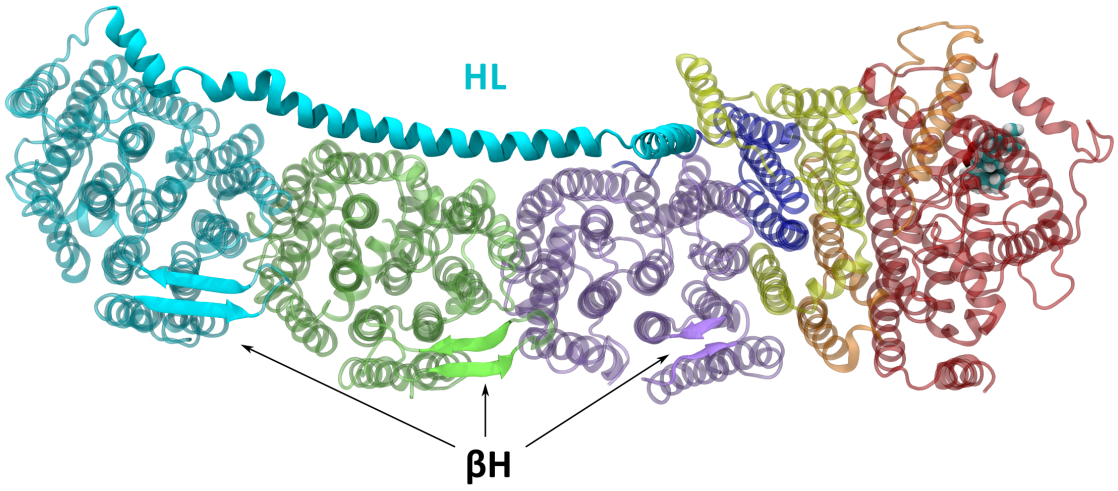
The bacterial complex I includes a smaller number of subunits than the mitochondrial one: a conserved core of 14 catalytic subunits. Prior to the evolution of mitochondria and eukaryotic cells, complex I included only the catalytic core subunits [39]. In addition to the 14 core subunits, mammalian complex I includes accessory subunits [35], bringing the number of subunits in the enzyme up to 45 [40]. The 14 core subunits are divided into two domains: the hydrophobic and the hydrophilic domains. The hydrophobic domain lies in the membrane, whereas the hydrophilic domain in the mitochondrial matrix. Seven of the core subunits are included in the hydrophobic domain and the rest in the hydrophilic domain. The seven subunits in the hydrophobic domain are mitochondrially encoded. The nomenclature of the core subunits differs among species. Table 2.1 summarises the nomenclature and shows the corresponding complex I core subunits in bacteria *E. coli* and *T. thermophilus*, in yeast *Y. lipolytica*, and in mammal *Bos taurus*.

Table 2.1 *Nomenclature of complex I core subunits.*

E. coli	*T. thermophilus*	*Y. lipolytica*	*Bos taurus*
NuoL	Nqo12	NU5M	ND5
NuoM	Nqo13	NU4M	ND4
NuoN	Nqo14	NU2M	ND2
NuoK	Nqo11	NULM	ND4L
NuoJ	Nqo10	NU6M	ND6
NuoA	Nqo7	NU3M	ND3
NuoH	Nqo8	NU1M	ND1
NuoI	Nqo9	NUIM	TYKY
NuoB	Nqo6	NUKM	PSST
NuoC	Nqo5	NUGM	30kDa
NuoD	Nqo4	NUCM	49kDa
NuoG	Nqo3	NUAM	75kDa
NuoE	Nqo2	NUHM	24kDa
NuoF	Nqo1	NUBM	51kDa



(a) Complex I sideview.



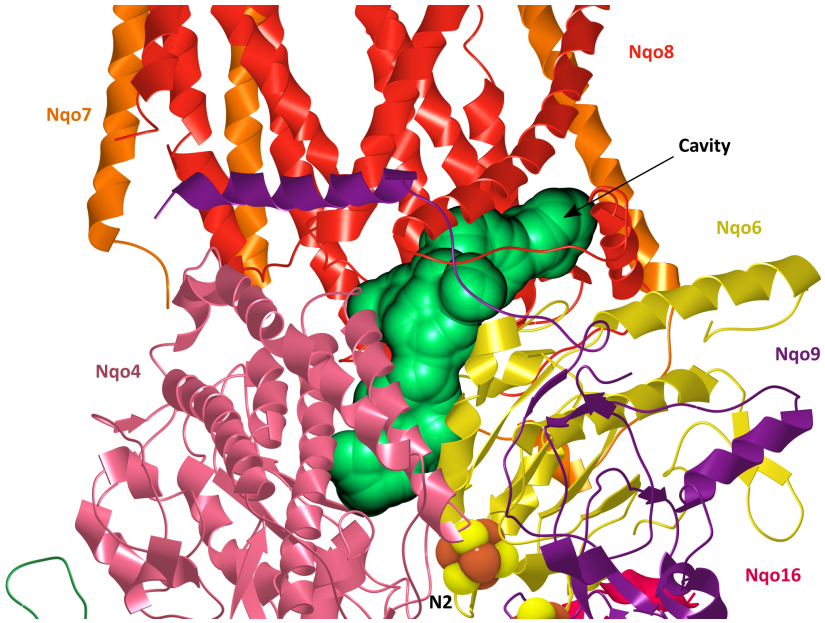
(b) Complex I top-down view of the membrane domain (P-side up).

Figure 2.5 *Structure of respiratory complex I from Thermus thermophilus. Upper panel shows the sideview, while the lower panel shows the top-down view from the P-side with a corresponding colour scheme. The hydrophilic domain is omitted in the lower image for clarity apart from the suggested position of the quinone.*

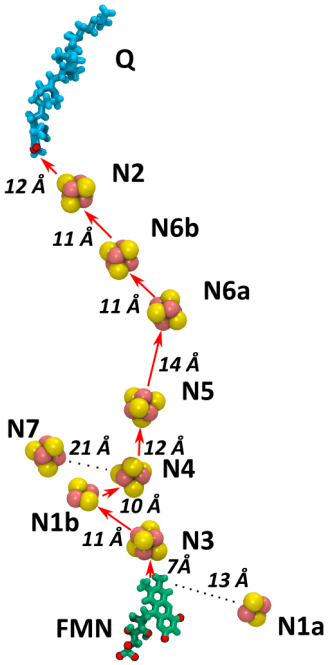
The hydrophilic domain consists of seven subunits residing in the mitochondrial matrix. In the crystal structure of complex I from *T. thermophilus*, there are however nine subunits in the hydrophilic arm. The extra ones, subunits Nqo15 and Nqo16, are organism specific, and the seven others constitute the conserved core [3]. All the redox-centers of complex I are in the hydrophilic domain, as well as the NADH binding site. NADH binding site is located in subunit Nqo1 in the solvent-exposed cavity, where FMN is non-covalently bonded [41]. Nearly 100 $\AA$ from the NADH binding site, in a cavity formed by subunits Nqo4, Nqo6, and Nqo8, lies the quinone binding site. These two sites are connected via a chain of redox centers, the iron-sulphur clusters. In *T. thermophilus* structure there are nine iron-sulphur clusters, of which seven have been suggested to form the electron transfer path. The seven iron-sulphur clusters are within a 14 $\AA$ distance from each other, which allows electron transfer via quantum mechanical tunnelling [42, 43]. Additionally, the final cluster, N2, is within 14 $\AA$ from the quinone binding site. Figure 2.6b shows the iron-sulphur clusters and the suggested electron transfer path. Figure 2.6a shows the cavity in complex I *T. thermophilus* structure in the junction of the hydrophilic and hydrophobic domains. The cavity was computed with the CAVER software [44, 45]. Quinone was fitted into this cavity in the model systems. A more detailed view of the residues surrounding the cavity is presented in Figure 2.6c.

The conserved core of the membrane arm consists of seven subunits, NuoL/ M/ N/ K/ J/ A/ H in *E. coli* and Nqo12/ 13/ 14/ 11/ 10/ 7/ 8 in *T. thermophilus*, respectively. They are embedded in the membrane next to one another forming a slightly curved linear structure (see Figure 2.5b). The extremity of the membrane domain spans 180 $\AA$ from the junction of the two domains, which in these scales is remarkable. Altogether, the membrane subunits include over 60 transmembrane helices, but surprisingly no redox centers, which was thought to be the case for a long time [34]. The observation that the membrane arm lacks redox centers had a major impact in understanding the function of complex I, because new theories regarding redox-coupled proton pumping had to be developed. To this date, the coupling mechanism of electron transfer and proton pumping has remained elusive.

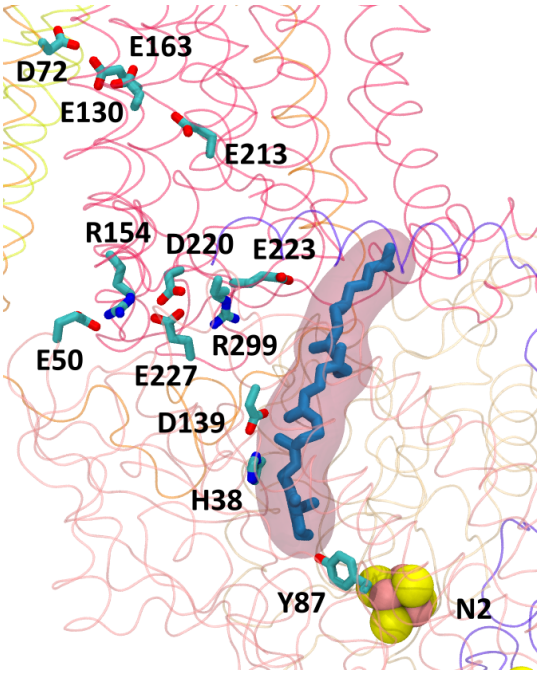
Subunit Nqo8/NuoH is located right on top of the hydrophilic domain, and its crystal structure was the last of the core subunits to be resolved. It forms the connection between the hydrophilic and hydrophobic domains, and additionally it is part of the quinone cavity. The helices in Nqo8 are tilted up to 40 degrees with respect to the



(a) Small cavity formed by subunits in the junction of membrane and hydrophilic domains.



(b) Iron sulphur clusters.



(c) The residues that line the quinone cavity in the model systems.

Figure 2.6 *The quinone binding cavity, the route of electron transfer via the iron-sulphur clusters in the hydrophilic domain of complex I, and quinone placement in the simulation systems. The cavity was computed with the CAVER software, and quinone is placed based on that. Quinone placement is similar to earlier study [46]. The electron transfer route via iron-sulphur clusters is indicated with red arrows.*

membrane domain. The tilt is visible in Figure 2.5b. The location and tilted helices of Nqo8, and its role in quinone cavity support the theory of conformational changes as the putative coupling mechanism of quinone reduction and proton pumping.

Subunits Nqo7/10/11 (NuoA/J/K) form a helix bundle that separates Nqo8 from the three remaining membrane subunits. The helix bundle has been suggested to include a proton translocation channel based on the conserved residues inside the subunits [47, 48]. Structurally the helix bundle together with subunit Nqo8 forms a reminiscent structural fold to that of subunits Nqo12-14. The fourth proton translocation channel, needed to fulfil the $4H^+/2e^-$ stoichiometry, has been proposed to be located within this structure. The exact location is however not known.

Subunits Nqo12/13/14 (NuoL/M/N) are Na^+/H^+ antiporter-like subunits, and form the extremity of the membrane arm. Each subunit has 14 transmembrane helices connected via loops, and a hairpin structure formed by two β-sheets (βH). Subunit Nqo12 additionally carries a long amphipathic helix, HL, which begins from transmembrane helix TMH15, continues along the cytoplasmic side of the membrane and ends with TMH16 at the interface of subunits Nqo14/11/10. Both βH and HL have been suggested to play a stabilising role in the membrane arm [8]. In HL, the residues interacting with the bulk of protein are relatively well conserved, but the rest of the helix is not [8]. This supports the view of HL being a mechanistic, stabilising element. The 14 transmembrane helices form a structure, that is unique among the known membrane protein structures [34]. The helices form two groups: TMHs1-3 and TMH14 are less conserved in comparison to TMHs4-14. In fact, all organisms do not include TMHs1-3 in all subunits [8, 49, 50]. TMHs4-14 form a conserved core of the subunits and they hold a unique symmetry: if TMHs4-8 are rotated 180 degrees and shifted, they can be superimposed on TMHs9-13. Similar symmetries have been observed in other transporters as well [51], but the fold in complex I is novel as the position of the TMs is face-to-back instead of face-to-face (see Figure 2.7). Based on the symmetry, it has been suggested that the antiporter-like subunits in complex I evolved via gene duplication [51].

Subunits Nqo15 and Nqo16 are not a part of the conserved core subunits of complex I. They are organism specific subunits. Subunit Nqo16 has been suggested to be an assembly factor [4].

(a) Helices before superimposing.

(b) Superimposed helices.

(c) Superimposed helices shown separately.

Figure 2.7 *The unique symmetry of transmembrane helices 4-8 and 9-13. Helices 4-8 are shown in purple and helices 9-13 in yellow.* ***(a)*** *The position of the helices in the crystal structure. The rotation axis is indicated with a black line. The chart in the bottom left of the figure shows the initial positions of the helices.* ***(b)*** *Helices 4-8 and 9-13 superimposed on each other. They superimpose well. Notice that helices 4-8 are fixed to the crystal structure conformation, but helices 9-13 have been rotated 180 degrees about the axis shown in (a).* ***(c)*** *The superimposed helices pulled apart to show the position of helices 9-13 better. The chart in the bottom-left corner shows the position of the two helix bundles after superimposion. The symmetry is face-to-back as helices 9-13 were only rotated about the axis shown in (a), not about the normal of this axis.*

Complex I is a reversible proton pump. It couples the reduction of ubiquinone to proton pumping accross the inner mitochondrial membrane, thus creating a proton electrochemical gradient. The current consensus value is $4H^+/2e^-$. The reaction cycle begins at the extremity of the hydrophilic domain, where NADH is oxidised by FMN:

$$NADH \rightarrow NAD^+ + 2e^- + H^+.$$

The reduction of FMN consumes the two electrons and two protons:

$$FMN + 2H^+ + 2e^- \rightarrow FMNH_2.$$

This redox reaction is the first electron transfer event in the function of complex I, and occurs via hydride transfer mechanism [34, p. 7]. The next step is the transfer of two electrons utilising seven iron-sulphur clusters in the hydrophilic domain. The

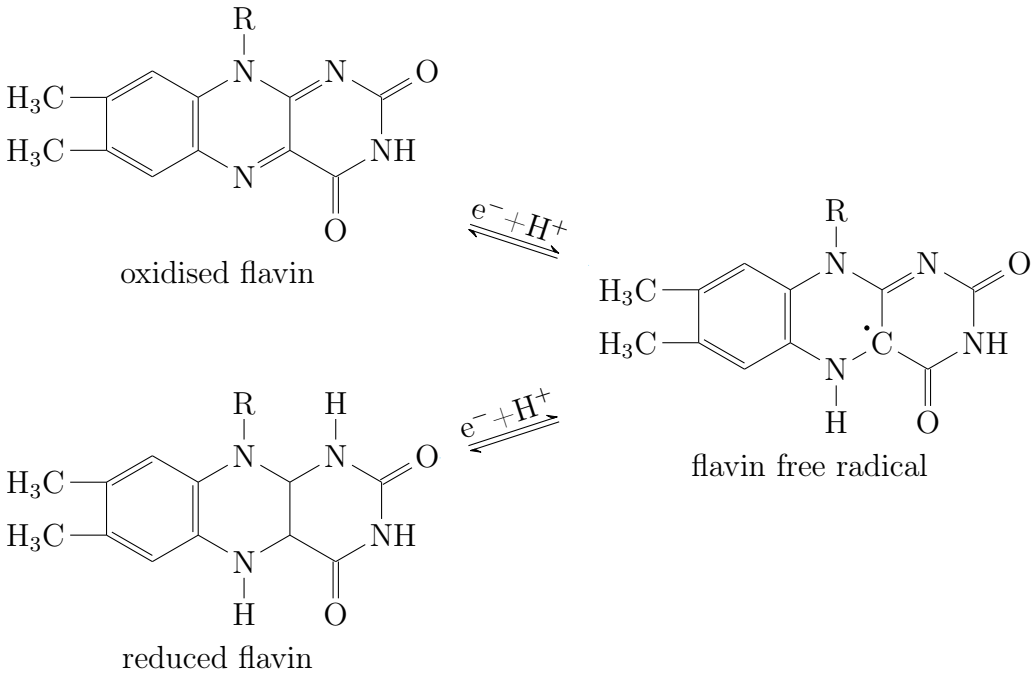


Figure 2.8 *The states of flavin. Part of the molecule has been replaced by R.*

main transfer route is

$$\text{N3} \rightarrow \text{N1b} \rightarrow \text{N4} \rightarrow \text{N5} \rightarrow \text{N6a} \rightarrow \text{N6b} \rightarrow \text{N7},$$

as is shown in Figure 2.6b. Iron-sulphur clusters are one-electron carriers, but NADH is a two-electron carrier. Two electrons are released from $FMNH_2$ in two steps, which includes the formation of a radical, flavin semiquinone, see Figure 2.8. After NAD^+ dissociates, the radical is exposed to solvent, hence becomes an efficient electron donor to oxygen, which leads to the production of reactive oxygen species [41, 52]. Iron-sulphur cluster N1a is not a part of the main electron transfer path, but it is conserved, which implies functional purpose. N1a can reduce the radicals and act as a protein-shielded electron storage reducing ROS production [34, pp. 7-8]. Cluster N7 is within 21 $\AA$ distance of the chain, and it is not conserved. It has been proposed to be an evolutionary remnant [3]. The final iron-sulphur cluster, called N2, lies approximately 12 $\AA$ from the proposed ubiquinone binding site allowing efficient electron transfer to quinone. N2 cluster is located in between two cysteines, which ligate the sulphur atoms of the cluster. When N2 is reduced, one of the cysteines dissociates and can take up a proton. It has been suggested that this in combination with the following subtle movement of helices would have a significant role in the function of complex I [5]. The electrons are transferred to ubiquinone one

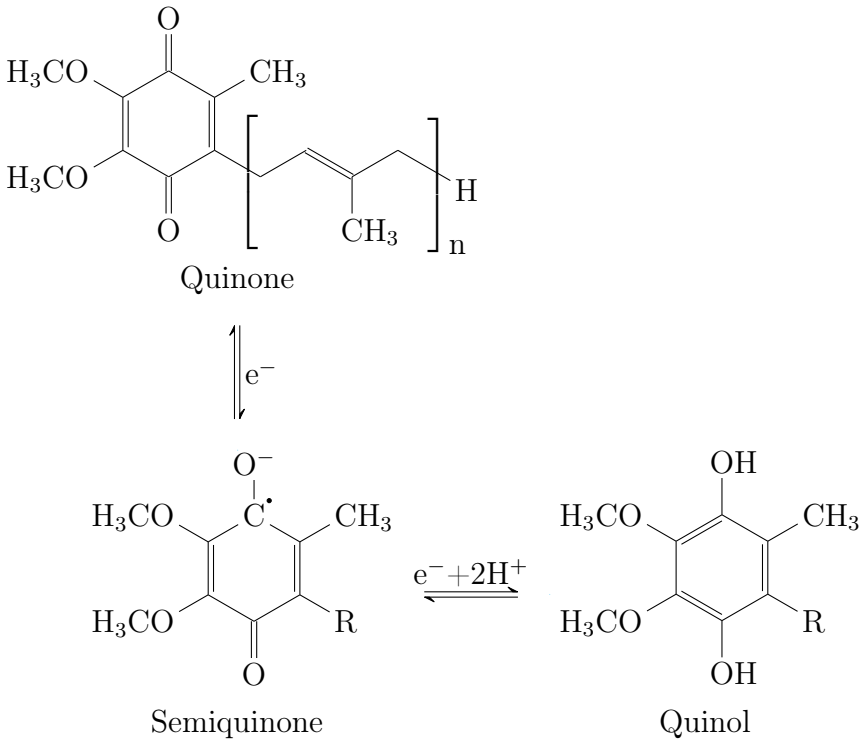


Figure 2.9 *Quinones and quinols. When n=10, the molecule is called ubiquinone.*

by one. First ubiquinone is turned to ubisemiquinone and then to Q^{2-}, which takes two protons from the surroundings and becomes charge neutral ubiquinol QH_2 [46]:

$$Q + 2e^- + 2H^+ \rightarrow QH_2.$$

The reduction of ubiquinone to ubiquinol releases free energy, which is utilized to perform proton pumping across the membrane. Details concerning ubiquinone redox reaction are shown in Figure 2.9.

Proton pumping in the membrane arm probably takes place via a Grotthuss-type mechanism [53–55]. Grotthuss mechanism requires a network of hydrogen bonds to function [54, 55]. Hydrogen bonds are long-range intermolecular forces that are much weaker than covalent bonds between atoms, but stronger than van der Waals forces [56, pp. 783-784]. Even though it is predicted that a Grotthuss-type mechanism would play a role in proton pumping, the details of this process are elusive. As the energetic cost of transferring a proton through membrane is too high, enzymes catalyse the transport, which suggests that proton channels must be present in the enzymes. Therefore, it is reasonable to study the water dynamics of complex I in or-

der to shed light on the possible proton pumping routes through the membrane arm.

As complex I pumps protons across the inner mitochondrial membrane against an electrochemical gradient, the protons would naturally flow back to the N-side and the gradient would equilibrate unless there are gating mechanisms. The electrochemical gradient is crucial to the function of complex V, which creates ATP from ADP and one phosphate. Regulating the proton flow is mandatory, but the gating mechanism in complex I is also elusive.

The current consensus value for proton pumping of complex I is four protons per two electrons. There are three homologous subunits in the membrane arm and their similarity to other proton pumps has led to the conclusion that these three subunits pump protons one per each subunits. The rest of the subunits, Nqo8, Nqo7, Nqo10, Nqo11, together form a structure similar to that of the three homologous subunits. It has been suggested that the fourth proton pathway would lie in these subunits. Each one of the proton translocation pathways is tentative, and as proton pumping is difficult to measure experimentally, computational methods have been adapted to study the proton pumping in the membrane. Only two earlier molecular dynamics studies related to the proton pumping pathways have been published [53, 57]. Both of these studies are based on the partial crystal structure of complex I from *E. coli*. Tan *et al.* performed MD simulations on two different protonation states of the membrane arm and observed a remarkable difference in hydration [57] between the two states. However, the chosen protonation states of residues were not argued with actual pKa calculations of the residues, but followed the lines of an earlier structural study [8]. Nevertheless, their results support that changing the protonation states of chosen residues would be connected to gating. Kaila *et al.* showed that interchanging symmetries of the residues leads to differences in hydration [53]. This study also supports the protonation state dependent coupling mechanism.

2.4 Motivation and research question

As there are only two earlier computational studies related to proton pumping in complex I, and both using an incomplete structure of the protein, the need for further investigation is apparent. The entire crystal structure of complex I has not yet been used for MD simulation studies regarding proton pumping mechanism. The aim of this study is to find the residues that participate in proton pumping pathways

in the membrane arm. Results of the three antiporter-like subunits are included in this work. The overall question regarding this study is related to the coupling mechanism of electron transfer and proton pumping. The observed results shed light on the connection between antiporter-like subunits, which is crucial in understanding the full mechanism of the entire cascade of events complex I performs.

Complex I is one of the main producers of reactive oxygen species (ROS) in mitochondria [58]. ROS can cause mitochondrial DNA damage, and they have been linked to Parkinson's disease [59] and ageing [60]. Other neurodegenerative diseases, such as Alzheimer's, have been connected to mutations in complex I [61–64]. Understanding the mechanism of complex I will enable studying how mutations cause dysfunctions and develop cures and medication to the diseases.

A computational approach, molecular dynamics simulations, was chosen to study complex I structure and function in atomic detail. Chapter 3 will introduce the background of the used methodology.

3. MOLECULAR DYNAMICS SIMULATIONS

This Chapter will introduce the physical and mathematical background of molecular dynamics (MD) simulations. The simulation software NAMD was used in this work, therefore the methods and tools available therein will be discussed here in detail.

3.1 Background

Proteins and other biological molecules often have complex physiological functions. Experimental studies provide valuable information about biological matter, and new and more accurate methods are constantly developed. However, certain research questions cannot be approached experimentally as the currently available experimental methods would possibly fail to produce relevant results. These questions are often specific and may concern atomistic details about nanoscale phenomena in time, for example the transition of oxygen atoms in and out of the heme pocket in myoglobin [65]. Computational methods enable studies of molecules on the atomistic level, as well as studies of phenomena whose timescale is so short that experimental methods are not applicable. Computational methods include, for example, Monte Carlo simulations and MD simulations. MD is a computational method, which aims to mimic the time-dependent behaviour of molecules as realistically as possible by sampling the configuration space through integration of Newton's equations of motion. MD simulations should be chosen, when observables are measured as a function of time [66]. MD simulations also require support from experimental studies, as it is crucial to compare computational and experimental results to secure the accuracy of modelling and identify needs for improvement. Additionally, MD simulations require the initial coordinates of atoms. Biological systems often include proteins, whose initial 3D structures are determined experimentally. MD simulations require parameters for atoms to identify them from each other, and to provide means to study the behaviour of particles realistically. These parameters are included in force fields, and will be discussed below. Computational and experimental

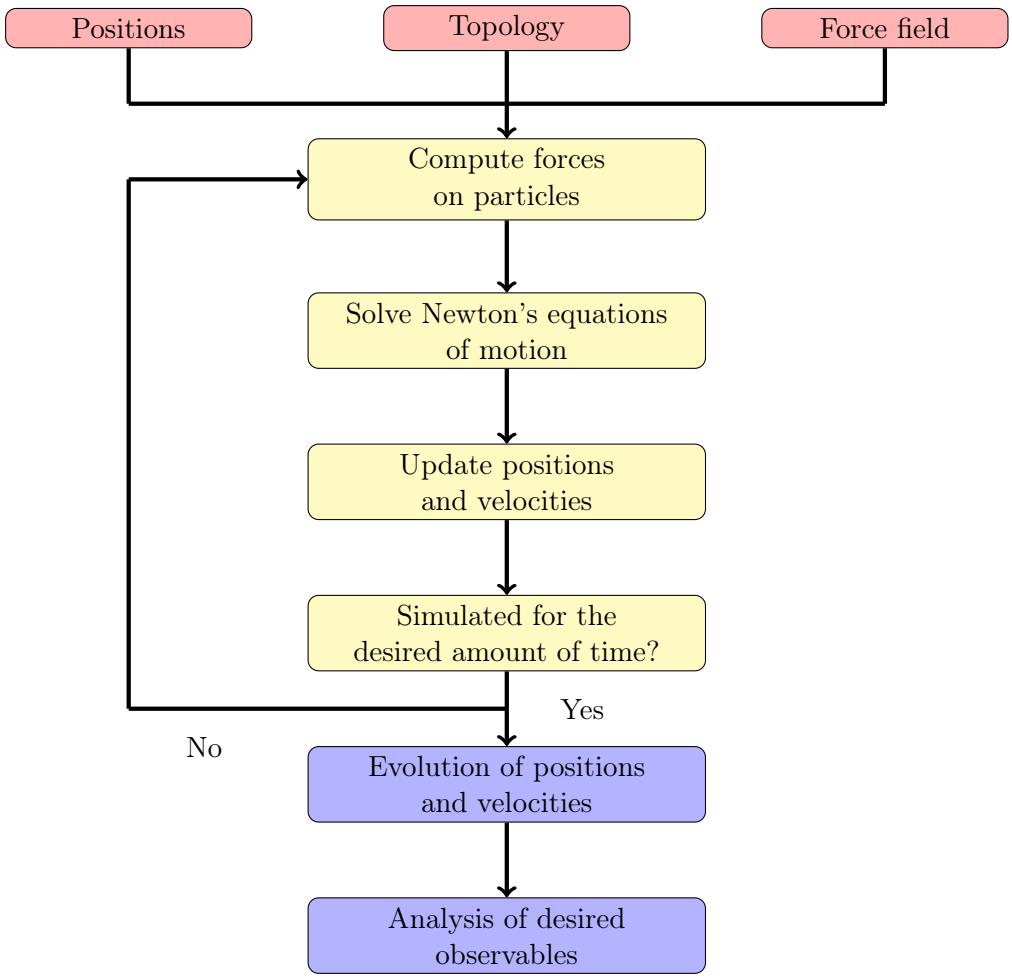


Figure 3.1 *General flow of classical all-atom MD simulation algorithms.*

studies support each other and one without the other would weaken their credibility.

MD simulations can be divided into different classes based on the nature of the model used to represent the results: all-atom or coarse grained.This study utilizes all-atom classical MD simulations. The particles in all atom classical MD simulations are atoms that interact with each other via bonded and non-bonded interactions, which govern the behaviour of the entire system. The flow of classical all-atom MD simulations is presented in Figure 3.1. Interactions and forces on atoms add to the

potential energy of the system, which is the key in computing a dynamic trajectory:

$$\begin{aligned} U_{total} &= U_{bonded} + U_{non-bonded} \\ &= \underbrace{U_{bond} + U_{angle} + U_{dihedral}}_{bonded} + \underbrace{U_{vdW} + U_{Coloumb}}_{non-bonded}. \end{aligned} \tag{3.1}$$

After the computation of the potential and the forces on particles, Newton's equations of motion are solved. The energy is updated. Depending on the chosen simulation ensemble, thermostats and barostats may be applied to conserve temperature and pressure, respectively. Thermostats and barostats will be discussed later.

The following sections will discuss the basic details of MD: interactions, force fields, minimization, periodic boundary conditions, integration, and temperature and pressure control.

3.2 Bonded interactions

Covalent bonds are strong interactions between atoms that form when orbitals of two atoms overlap, so that they both achieve a full valence shell [56, p. 621]. The type of overlapping orbitals determines the type of the bond. Mechanistically speaking, covalent bonds restrict the movement of atoms. In order to model covalent bonds in MD simulations, it is conventional to determine three internal coordinates for the system: bond, angle and dihedral. Internal coordinates differ from the usual Cartesian coordinates, because they are defined relative to other atoms. Internal coordinates are shown in Figure 3.2. Dihedral angle (ϕ in Figure 3.2) describes the torsion of two planes formed by four atoms. In Figure 3.2 these planes are ABC and BCD. Dihedral angles can be divided into proper dihedrals and improper dihedrals. Improper dihedrals restrict torsion in such a way that planar structures, such as aromatic rings, remain planar throughout the simulation.

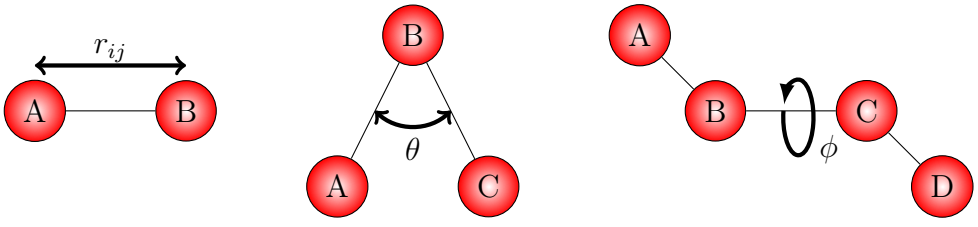


Figure 3.2 *Internal coordinates for bonded interactions.*

Potential energy is defined as a function of the internal coordinates, because it depends on the distances and angles between particles rather than the absolute coordinates in a three-dimensional space. In this study, the simulation program NAMD is used, hence the potential energy functions are described as they are defined in NAMD documentation. Bond stretching is modelled as two masses joined by a harmonic spring according to Hooke's law:

$$U_{bond} = \sum_{bonds\ ij} k_{f,ij}(r_{ij} - r_{f,ij})^2, \tag{3.2}$$

where $k_{f,ij}$ is the force constant, r_{ij} is the distance between the two atoms i and j, and $r_{f,ij}$ is its reference value. Angle bending is also described with a harmonic potential:

$$U_{angle} = \sum_{angles\ i} k_{f,\theta}(\theta_i - \theta_{f,i}), \tag{3.3}$$

where $k_{f,\theta}$ is the force constant, θ_i is the angle between three atoms, and $\theta_{f,i}$ is its reference value. Torsion angle bending, or dihedral angle, is described with

$$U_{dihedral} = \sum_{dihedrals\ i} \begin{cases} k_{f,\phi}[1 + \cos(n_i\phi_i - \phi_{f,i})], n_i \neq 0 \\ k_{f,\phi}(0_i - \phi_{f,i})^2, n_i = 0 \end{cases}, \tag{3.4}$$

where $k_{f,\phi}$ is the force constant for the torsion angle bending potential, ϕ_i is the torsion angle as defined in Figure 3.2, and $\phi_{f,i}$ is the reference value. n_i describes the number of minima, when the bond is rotated 360 degrees; multiplicity of the bond.

Covalent bonds in MD simulations are described with harmonic potentials treating atoms and bonds as masses joined by springs. The approach provides a simple and computationally beneficial model for bonded interactions. The force constants and the reference values in the interaction potentials are included in force fields, which will be discussed below.

3.3 Non-bonded interactions

Non-bonded interactions include van der Waals and electrostatic interactions. Van der Waals interactions are described with the Lennard-Jones 6-12 potential and

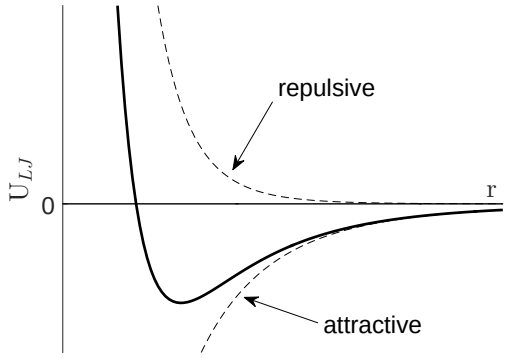


Figure 3.3 *Lennard-Jones potential. Lennard-Jones potential consists of an attractive and a repulsive term.*

electrostatics are handled with the Coulomb potential. In both cases, the potential energy is measured as a function of distance between two atoms. It would be computationally too costly to measure the non-bonded interactions of every atom pair in the system, thus cutoffs are commonly used. However, due to the long-ranged nature of electrostatic interactions, pure cutoff techniques are not recommended, but special approaches such as PME [67] give a much more accurate description. In NAMD, the non-bonded potential energy is defined in terms of the van der Waals potential energy, U_{vdW}, and the electrostatic potential energy, $U_{Coulomb}$:

$$U_{vdW} = \sum_i \sum_{j>i} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right], \tag{3.5}$$

$$U_{Coulomb} = \sum_i \sum_{j>i} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}, \tag{3.6}$$

where ϵ_{ij} and σ_{ij} are dependent on the type of atoms, and they describe the potential well depth and the distance at which the interaction is zero, respectively. The charge of the atom is marked with q, and ϵ_0 is the dielectric constant in vacuum. In both equations r_{ij} is the distance between atoms. Lennard-Jones potential consists of the repulsive term, being r^{-12}-dependent, and the attractive term, being r^{-6}-dependent. Figure 3.3 shows how the two terms are related, how they give a unique potential energy profile together, and how the potential weakens as the distance between interacting particles increases.

3.4 Force fields

The preceding Sections 3.2 and 3.3 discuss the different types of interactions employed in MD simulations. Each type of interaction requires predefined parameters, which are characteristic for atom combinations and atom types. Essentially, force fields are collections of these parameters. The choice of force field is one of the most important steps in simulation system preparation as it will in practise determine the accuracy of modelling.

Properties of atoms do not only depend on the element, but also how they are covalently bonded. Thus, force fields include several atom types for each element. For instance, benzene ring carbons require different parameters than acyl chain carbons.

Force field parameters in classical force fields are determined by quantum chemical calculations and experimental data and are optimised to give the best results. The determination of force field parameters requires specific expertise and takes a great amount of time, hence most MD studies utilise the available force fields. There are several force fields, such as CHARMM [68], AMBER [69] and OPLS [70] to mention a few. Additionally, force fields usually have more than one available version with small differences.

3.5 Minimization

Before the beginning of an actual MD simulation, the system is minimised to find a low-energy conformation. Minimization is a process during which an optimization algorithm is applied to the system aiming to find the minima of the nonlinear potential energy function. The potential functions in the standard force fields are generally multivariate, nonlinear and continuous, which limits the number of applicable optimization algorithms [71]. Additionally, the requirements for optimization algorithms for biological systems are high as they must be computationally low cost and require memory moderately even for large-scale systems, but at the same time they must find the local minima efficiently from any point without getting caught to saddle points or local maxima. Effective large-scale multivariate optimization algorithms often used in simulation programs include for instance Quasi-Newton, Conjugate Gradient (CG) and Truncated-Newton. The simulation software used in this study uses the CG algorithm, hence it will be explained in detail.

As the basic CG algorithm iteratively finds the minima of linear systems [72], it is combined with line-search methods to make CG appropriate to nonlinear systems as well. For readability, the minimization problem will be denoted as

$$\min f(\mathbf{x}),\ \mathbf{x} \in \mathbb{R}^n, \tag{3.7}$$

where f is the potential energy function and x corresponds to the configuration. The CG algorithm to solve this energy minimization starts from the initial configuration $\mathbf{x}_0$ and goes on for $k = 0, 1, 2, \ldots$ [73, pp. 318-319]:

1. Check the current configuration, $\mathbf{x}_k$, for convergence.

2. Determine the search vector $\mathbf{p}_k$, which implies the descent direction. This step depends on the chosen method, and in CG it is defined as

$$\mathbf{p}_k = -\mathbf{g}_k + \beta_k \mathbf{p}_{k-1}, \tag{3.8}$$

   where β_k is the conjugate gradient parameter, and $\mathbf{g}_k$ is the gradient defined as

$$\mathbf{g}_k = \frac{\partial f(\mathbf{x}_k)}{\partial \mathbf{x}_k}. \tag{3.9}$$

   The determination of β_k is dependent on the simulation software, and several version and modifications exist [73, p. 333], for example the Polak-Ribiére algorithm:

$$\beta_{k+1} = \mathbf{g}_{k+1}^T \mathbf{y}_k / \mathbf{g}_k^T \mathbf{g}_k, \tag{3.10}$$

   where $\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$.

3. From step 2 the direction for the next iteration is obtained, and in step 3 the step length λ_k is determined. This is executed by one dimensional line search. The new configuration $\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k \mathbf{p}_k$ and its corresponding gradient has to satisfy two conditions:

$$f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) + \alpha \lambda_k \mathbf{g}_k^T \mathbf{p}_k \tag{3.11}$$

$$|\mathbf{g}_{k+1}^T| \leq \beta |\mathbf{g}_k^T|, \tag{3.12}$$

   where $0 \leq \alpha \leq \beta \leq 1$.

4. Set the new configuration $\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k \mathbf{p}_k$ and return to step 1.

3.6 Periodic boundary conditions

Due to limited computing resources, model systems cannot be of infinite size, and as a result are prone to edge effects. To overcome this issue, periodic boundary conditions (PBC) are applied [74, p. 317]. In PBC, atoms feel the forces as if they were embedded in bulk fluid by replicating the unit cell in all directions [74,75] filling all of space. The most used unit cell shapes are a cube and a truncated octahedron [74, p. 319], but several other shapes also exist. Essentially, any shape could be used as long as it can fill all of space with translations of unit cell components in all dimensions. Naturally, the choice of the shape depends on the shape of the model system. Model systems in this work consist of an L-shaped protein embedded in a membrane, hence a rectangular unit cell was chosen.

3.7 Integrator

MD simulations require computation of the Newton's equations of motion for every timestep. The usage of thermostats and barostats will modify the Newton's equations slightly, as will be seen in Section 3.8.

The potential energy of the system, U_{total}, is used to obtain the forces acting on each particle,

$$\mathbf{F}_i = -\nabla U_{total}. \tag{3.13}$$

After the forces have been calculated, the motion of particles can be solved from Newton's equation

$$\mathbf{F}_i = m_i \frac{\partial^2 \mathbf{x}_i}{\partial t^2}, \tag{3.14}$$

where m_i is the mass and $\mathbf{x}_i$ the position of the particle i. Continuous potential function couples the motions of particles. Therefore, simulations on biological systems are chaotic many-body problems; hence, the solution to the differential equation is not obtained analytically, but by numerical integration. The simulation program NAMD utilises the velocity Verlet method, which is a second order numerical integrator. Velocity Verlet [76] is a modification of the Verlet method [77]. Verlet method -based integrators are used in other simulation software as well, for instance leapfrog integration employed in the simulation software GROMACS [78]. Other finite difference methods apart from the Verlet method also exist, but all of them are based on the assumption that positions, velocities and accelerations etc. can be

approximated with Taylor series expansions:

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \Delta t\mathbf{v}(t) + \frac{1}{2}\Delta t^2\mathbf{a}(t) + \frac{1}{6}\Delta t^3\mathbf{b}(t) + \frac{1}{24}\Delta t^4\mathbf{c}(t) + \cdots \tag{3.15}$$

$$\mathbf{v}(t+\Delta t) = \mathbf{v}(t) + \Delta t\mathbf{a}(t) + \frac{1}{2}\Delta t^2\mathbf{b}(t) + \frac{1}{6}\Delta t^3\mathbf{c}(t) + \cdots \tag{3.16}$$

$$\mathbf{a}(t+\Delta t) = \mathbf{a}(t) + \Delta t\mathbf{b}(t) + \frac{1}{2}\Delta t^2\mathbf{c}(t) + \cdots \tag{3.17}$$

$$\mathbf{b}(t+\Delta t) = \mathbf{b}(t) + \Delta t\mathbf{c}(t) + \cdots, \tag{3.18}$$

where $\mathbf{r}$ is the position, $\mathbf{v}$ the velocity, which is the first derivative of position, and $\mathbf{a}$ the acceleration, which is the second derivative of position, $\mathbf{b}$ and $\mathbf{c}$ are the first and second derivatives of acceleration, respectively. In order to calculate positions at time $t + \Delta t$, the Verlet algorithm uses the positions and accelerations at time t, and the positions at time $t - \Delta t$. Δt stands for the time step of integration, which in biological simulations is often in the range of a few femtoseconds. Positions at time $t + \Delta t$ and $t - \Delta t$ can be written as

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \Delta t\mathbf{v}(t) + \frac{1}{2}\Delta t^2\mathbf{a}(t) + \cdots \tag{3.19}$$

$$\mathbf{r}(t-\Delta t) = \mathbf{r}(t) - \Delta t\mathbf{v}(t) + \frac{1}{2}\Delta t^2\mathbf{a}(t) + \cdots \tag{3.20}$$

When Equations (3.19) and (3.20) are added together, the result is

$$\mathbf{r}(t+\Delta t) \approx 2\mathbf{r}(t) - \mathbf{r}(t-\Delta t) + \Delta t^2\mathbf{a}(t) \tag{3.21}$$

Velocities are not obtained directly from the Verlet method, but they must be calculated separately

$$\mathbf{v}(t) = \frac{\mathbf{r}(t+\Delta t) - \mathbf{r}(t-\Delta t)}{2\Delta t} \tag{3.22}$$

or

$$\mathbf{v}(t+\frac{1}{2}\Delta t) = \frac{\mathbf{r}(t+\Delta t) - \mathbf{r}(t)}{\Delta t} \tag{3.23}$$

Equations (3.21) and (3.22)/(3.23) constitute the basis of the Verlet algorithm. In order to solve the acceleration $\mathbf{a}$, forces on particles have to be computed from the potential function. In the Verlet method, a very small term $\Delta t^2\mathbf{a}(t)$ is added to two much bigger terms in Equation (3.21), which may lead to a small loss of precision.

The velocity Verlet method uses the current position and velocity to solve for the

position and velocity of the next step using the forces $\mathbf{F}_i$ [75]:

$$\begin{aligned}
\mathbf{v}_{n+1/2} &= \mathbf{v}_n + M^{-1}\mathbf{F}_n \cdot \Delta t/2 = \mathbf{a}_n \cdot \Delta t/2 \\
\mathbf{r}_{n+1} &= \mathbf{r}_n + \mathbf{v}_{n+1/2}\Delta t \\
\mathbf{F}_{n+1} &= \mathbf{F}(r_{n+1}) \\
\mathbf{v}_{n+1} &= \mathbf{v}_{n+1/2} + M^{-1}\mathbf{F}_n \cdot \Delta t/2 = \mathbf{v}_{n+1/2} + \mathbf{a}_n \cdot \Delta t/2,
\end{aligned}$$

where $\mathbf{r}$ is the position of particles, M is mass matrix and Δt is timestep. The great advantage of the Verlet method is that force evaluation per timestep is required only once. Force evaluation is the most time consuming step in MD, thus integration methods that require several force evaluations per time step, such as Runge-Kutta type of methods, are not efficient [75]. The global error of the velocity Verlet is proportional to Δt^2, which is in the order of magnitude 10^{-30}, when the time step is a few femtoseconds [75]. The Verlet method and its variations are popular among biomolecular simulation software, because other symplectic integrators are less accurate or not as practical.

3.8 Temperature and pressure coupling

In order to maintain biologically relevant conditions in MD simulations, macroscopic quantities must be maintained. In a microcanonical ensemble (NVE) the number of particles, volume of the system, and total energy are held constant. The canonical ensemble (NVT) corresponds to a thermostat when the number of particles, volume of the system and temperature are conserved. The isothermal-isobaric ensemble (NPT) corresponds to a barostat in addition to the thermostat, where the number of particles, pressure and temperature are constant. Experiments are often carried out in either NVT or NPT conditions; hence, these ensembles are often applied to MD simulations. The integrator produces the correct ensemble distribution by coupling the system to a reservoir, which is achieved by adjusting the Newton's equations of motion slightly. Several algorithms for temperature and pressure coupling are available, each with their own advantages. Temperature coupling algorithms include, for instance, the Nosé-Hoover thermostat [79, 80], the Berendsen thermostat [81], the Andersen thermostat [82] and Langevin dynamics [83]. Pressure coupling algorithms include, for instance, the Berendsen barostat [81] and the Parrinello-Rahman barostat [84, 85].

The simulations in this study are performed in NPT conditions, and Langevin dynamics was used for temperature control, and the Nosé-Hoover Langevin piston for pressure control. The Nosé-Hoover Langevin piston pressure control is a modified Nosé-Hoover barostat implemented in the simulation program NAMD. This pressure control algorithm combines the Nosé-Hoover barostat to Langevin dynamics by controlling the fluctuations in the barostat with Langevin dynamics.

Langevin dynamics is based on the continuous Langevin equation [73, p. 435]

$$M\ddot{X}(t) = -\Delta E(X(t)) - \gamma M\dot{X}(t) + \mathbf{R}(t), \tag{3.24}$$

where M is mass matrix, $X(t)$ is the coordinate matrix, $\Delta E(X(t))$ is the change in potential energy, γ is the damping coefficient, and $\mathbf{R}(t)$ is a random-force vector. The utilisation of Langevin dynamics requires small adjustments to the Verlet algorithm [73, p. 437].

3.9 Limitations of MD simulations

Molecular dynamics is an efficient computational method to study various types of systems. Biologically relevant topics are popular among the computational research, because computational methods shed light on phenomena out of reach of the currently available experimental techniques. However, computational methods do have their limitations.

One of the most crucial limitation in computational science is computing resources. Even with highly efficient simulation programs and algorithms, MD simulations require massive amounts of computation time. Computation time is limited and expensive, hence careful modelling is necessary to avoid flawed model systems and the need for recalculation. Even small mistakes in the system may accumulate in the course of the simulation and produce incorrect results.

Concerning the accuracy of simulations, force fields are a major limiting step in molecular dynamics simulations. The parameters in force fields are approximations, and the description of interactions achieved from them is limited in accuracy. However, the available force fields are well tested and they have been proven to produce reliable results for a variety of compounds. The parameters in force fields are employed by potential functions, which affect the behaviour of the system the most.

Covalent bonds are approximated with harmonic potentials, and the computation of non-bonded interactions utilises cutoffs.

Another major issue with computational methods is related to sampling. Due to computational resources, the time length and the size of systems are limited. Biological phenomena may occur in milliseconds to seconds. Studying such phenomena on atomistic scale with MD simulations is difficult without special techniques, such as accelerated dynamics. In addition, the systems cannot be of infinite size, hence small systems may be affected by finite size effects. PBCs aim to reduce them.

In MD simulations, the motions of particles are computed with classical mechanics; hence, quantum phenomena cannot be studied with MD. Even though the parametrization of force fields is performed carefully with quantum mechanical calculations, quantum phenomena in the simulation systems is excluded.

Performing MD simulations is challenging due to the aforementioned points. The basic theory and algorithms are well understood and tested, but the user has to be careful in choosing the force field and the algorithms for each system correctly. Additionally, MD simulations produce massive amounts of data to analyse, which requires special approaches for data handling and analysis.

4. METHODS

The crystal structure of the entire respiratory complex I has been resolved at 3.3 $\AA$ resolution, therefore it does not contain crystallographically resolved water molecules. In this work, a computational approach was chosen to study the water dynamics in the respiratory complex I.

This section presents the methods that were used to construct and analyse the model systems. Theoretical details of the computational methods were introduced in the previous chapter (Chapter 3).

4.1 Model systems and simulation details

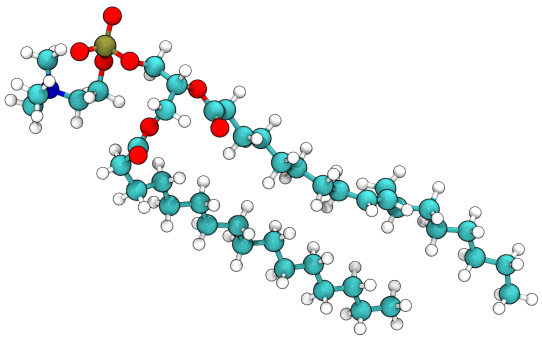


Figure 4.1 *1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid. Carbon is shown cyan, hydrogen white, oxygen red, phosphorus mustard and nitrogen blue.*

Two model systems were constructed and simulated for this study. The crystal structure of the entire respiratory complex I from *T. thermophilus* (PDB id 4HEA [4]) was embedded into a homogeneous 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid bilayer. The structure of POPC lipid is shown in Figure 4.1. Three short segments, residues 56-72 and 144-147 in Nqo3 and 65-69 in subunit Nqo6, are not resolved in the crystal structure. Therefore, they were modelled based on earlier

work [46], in which the program MODELLER [86] was used to model the unresolved fragments. Quinone was modelled as Q_6 either in the oxidized form (Q_{ox}) or as quinol (QH_2) [4]. Depending on the form of quinone in the system, protonation states of the nearby residues in subunit Nqo4 were determined [46]. The protein membrane system was solvated with explicit solvent using the TIP3 water model and ionised with 100 mM Na^+ and Cl^- ions in order to mimic physiological conditions. Due to the L-shape of complex I, the number of atoms in the unit cell was high (823,701-823,705). The overall view of the model systems is shown in Figure 4.2 and the details in Table 4.1.

Table 4.1 *Model systems.*

Model system	Quinone	H38	Y87	D139	Length (ns)
1	Q_{ox}	+1	0	-1	498
2	QH_2	0	-1	-1	501

All-atom molecular dynamics simulations were performed with the program NAMD2 [75] using the CHARMM force field for protein, lipids and the redox centers [68,87]. Force field parameters for the redox centers were obtained from references [88–91]. Parameters not found in the previous references were obtained from earlier work [46]. The timestep for integration was 1 fs, pressure 1 atm and temperature 310 K, which were controlled with the Langevin barostat and thermostat. Particle Mesh Ewald (PME) was used to calculate electrostatics [67]. Total simulation time for the two systems analysed in this study was approximately 1 μs, about 500 ns each.

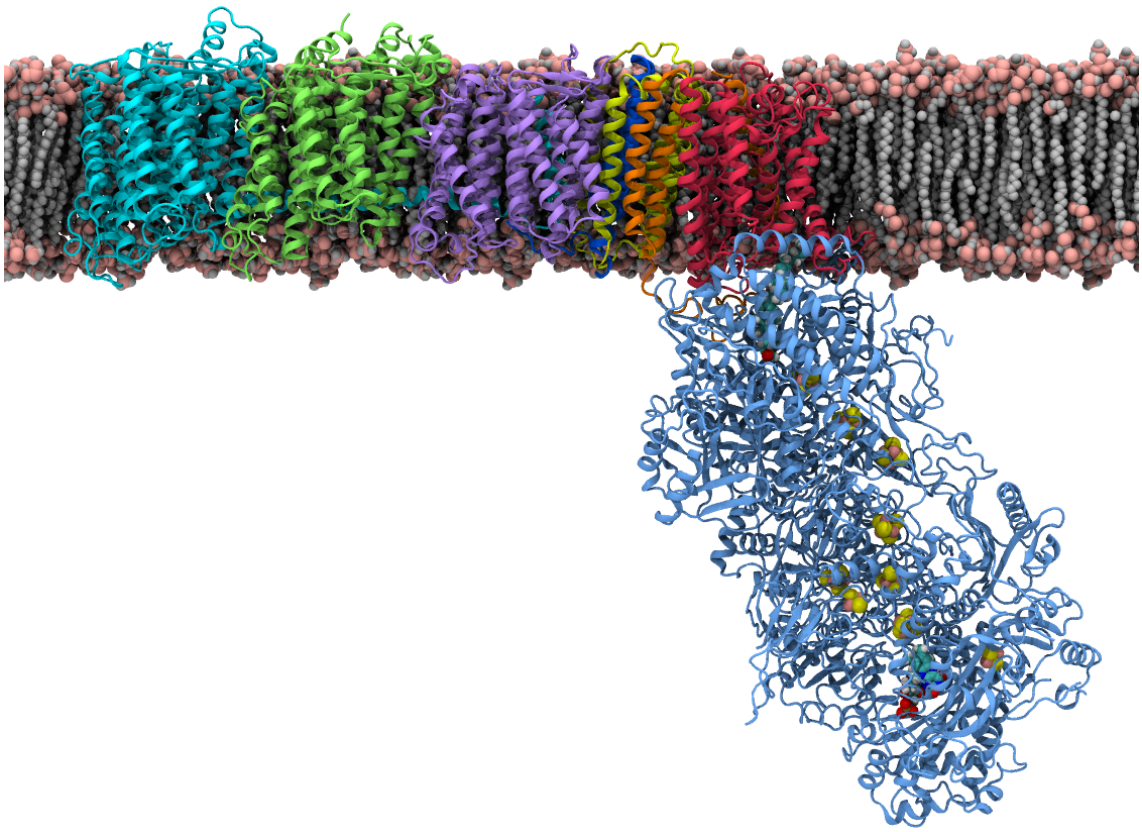


Figure 4.2 *Model system. Solvent and ions are omitted for clarity. Subunits Nqo12-14 are shown in cyan, green and purple ribbons, subunit Nqo7 and Nqo8 in orange and red ribbons, subunits Nqo10 and Nqo11 in yellow and dark blue ribbons, repectively. All of the hydrophilic domain subunits are shown in blue ribbons, with iron-sulphur clusters, quinone, and FMN as spheres. The lipid tails are shown as grey spheres, and lipid headgroup atoms shown as pink spheres.*

4.2 Analysis tools

The simulation data and protein structures were analysed with the Visual Molecular Dynamics (VMD) [92] software. Figures were prepared and rendered using VMD and further processed with Inkscape and GNU Image Manipulation Program (GIMP) [93]. Plots were prepared with Matlab [94] and LaTeX.

4.2.1 Alignment

Sequence alignment was performed with the program Clustal [95, 96]. Multiple Sequence Alignments (MSA) are routinely used to identify conserved residues in

protein sequences obtained from several species or to study the evolutionary history of a protein [97]. Sequence alignment can be either purely sequential or structure-based, in which gap penalties are added to regions with secondary structures, such as helix. Highly conserved residues in amino acid sequences may be linked to functionally and mechanistically important regions in the protein structure.

Structural alignment was performed with the VMD plugin Multiseq [98], which utilises Structural Alignment of Multiple Proteins (STAMP) program [99] for aligning proteins. The aim of the STAMP algorithm is to minimise the alpha carbon distance between aligned residues of proteins by adding appropriate translations and rotations to the protein structure.

Similarities between structures can be compared by measuring Q_{res}. Q_{res}, derived from the Q parameter for protein folding, is a value between 0 and 1, and can be used to identify structural similarities in aligned proteins for each residue [98]. Q_{res} value 1 corresponds to perfect structure alignment, whereas 0 corresponds to no structure alignment whatsoever. While Q computes native contacts, Q_{res} computes and compares alpha carbon to alpha carbon distances between the residue under investigation and all other residues. The nearest neighbouring residues are however omitted. These values are then compared to the corresponding values in the set of aligned proteins. Q_{res} values are computed by

$$Q_{res}^{(i,n)} = N \sum_{m \neq n}^{proteins} \sum_{j \neq i-1,i,i+1}^{residues} \exp\left[-\frac{(r_{ij}^{(n)} - r_{i'j'}^{(m)})^2}{2\sigma_{ij}^2} \right], \tag{4.1}$$

where N is a normalization factor, which scales the values of Q_{res} to be between 0 and 1, $r_{ij}^{(n)}$ is the distance between alpha carbons between residues i and j in protein n and $r_{i'j'}^{(m)}$ is the same distance between the corresponding residues in protein m. The normalization factor is defined as

$$N = \frac{1}{(N_{seq} - 1)(N_{res} - k)}, \tag{4.2}$$

where N_{seq} is the number of proteins and N_{res} is the number of residues in protein n. The value of k is 3 unless residue i is in N- or C-terminus. In that case the value of k is 2, because a terminating residue has one nearest neighbouring residue instead of two. The variance term, σ_{ij}^2, in Equation (4.1) is computed from the sequence

separation of residues i and j

$$\sigma_{ij}^2 = |i-j|^{0.15} \tag{4.3}$$

The excluded neighbours in Q_{res} computation are the previous residue in sequence and the next one with respect to the one under inspection, as can be seen in Equation (4.1).

4.2.2 Stability measurements

Root-mean square deviation (RMSD) measures the difference between the structures at time t_2 and the reference structure at time t_1 [100]. RMSD is calculated as

$$\mathrm{RMSD}(t_1, t_2) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} ||\mathbf{x}_i(t_2) - \mathbf{x}_i(t_1)||^2}, \tag{4.4}$$

where N is the number of atoms in the system, $\mathbf{x}_i(t_2)$ is the position of atom i at time t_2, and $\mathbf{x}_i(t_1)$ is the reference position of atom i [101]. RMSD analysis is a simplistic analysis to see stabilization of the system.

The deuterium order parameter (S_{CD}) is defined as

$$\mathrm{S}_{CD} = \frac{3}{2}\left\langle \cos^2 \omega \right\rangle - \frac{1}{2}, \tag{4.5}$$

where ω is the angle between the carbon-hydrogen bond and the membrane normal. The value is averaged over both CH_2-group C-H bonds, over all the lipids and in time [102]. S_{CD} measures how ordered the lipid tails are, and thus it is often used to estimate the phase of the membrane. The value of S_{CD} is calculated to each CH_2-group, and it is between -0.5 and 1.0. The more ordered the lipid tails are, the higher the value of S_{CD} is.

4.2.3 Water channel analysis

Water wires in subunits Nqo12-14 were analysed with a combination of methods. Visualisation was extensively used to choose a tentative set of amino acid residues participating in water wire formation. The tentative set was used to define water

molecules inside the subunits to be passed to hydrogen bond analysis tool as implemented in VMD [92] . Hydrogen bonds were analysed between the water molecules inside the subunits and all of the residues of the subunits with the hydrogen bond criteria; within a 3.5 $\AA$ donor-acceptor distance and 30 degree bond angle. The results are shown in Tables S.4, S.5, and S.6. The tables list the residues obtained in model systems 1 and 2 with a maximum hydrogen bond occupancy higher than 10 %. Hydrogen bond analysis resulted in a remarkable number of hydrophobic residues, which border the water wires. The aim was to assign residues, which actually participate in water wire formation, so the hydrophobic residues were excluded. The remaining set of residues was parsed based on residue conservation: non-conserved residues were removed. The transient connections to P-bulk and N-bulk were obtained visually, and the results were added to the earlier set. The resulting set of residues consists of highly conserved hydrophilic residues, which participate in water wire formation, and thus also in proton pumping through the membrane domain.

Mutation data for complex I is available from several earlier studies [47, 48, 103–111]. The residue numbering in these studies follows *Escherichia coli* nomenclature, the residue numbering was translated to correspond *Thermus thermophilus* convention using sequence alignment table for complex I subunits [8]. Proton pumping activity is difficult to measure; hence, the percentages of activities are approximate. Mutation data was cross-checked with the residues assigned to water channel formation to identify functionally important residues, and validate the choice of residues coordinating water channel formation.

4.2.4 Helix HL and water molecules

The behaviour of water molecules around helix HL was analysed by measuring how many water molecules pass by helix HL to the small space left between the protein and the membrane. The number of water molecules around a cylinder of 5 $\AA$ radius around the long horizontal helix was measured. A 2D projection of the Cα atoms of the helix and the water molecules was made. By least squares fitting, a line through the Cα atom projection was fitted and the percentage of water molecules above helix HL measured as a function of simulation time. Figure 4.3 shows an example of the projection along with the line fit.

In practise, this analysis tool will measure the percentage of water molecules above

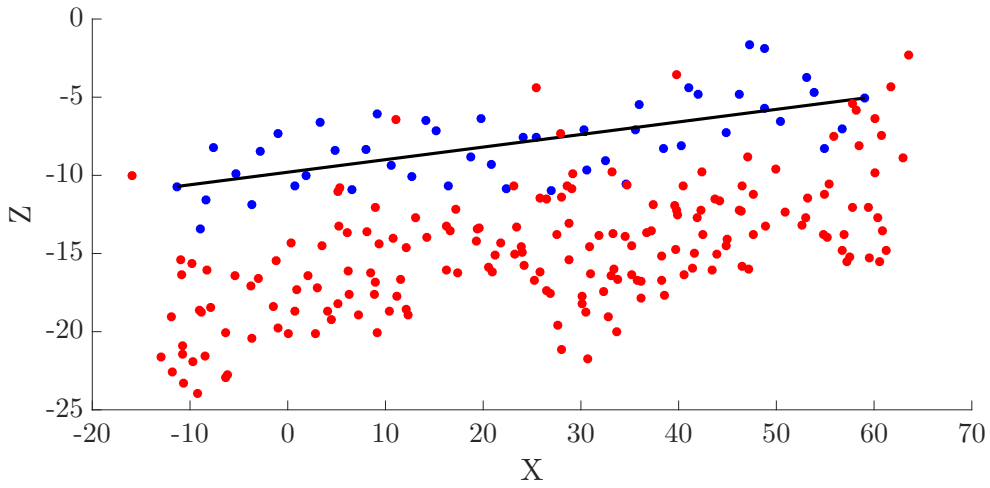


Figure 4.3 *Example of the 2D projection of coordinates. Protein Cα atoms are shown as blue circles, water molecules as red circles and the least-squares fitted line through the Cα coordinates with black. The example plot is from model system 1.*

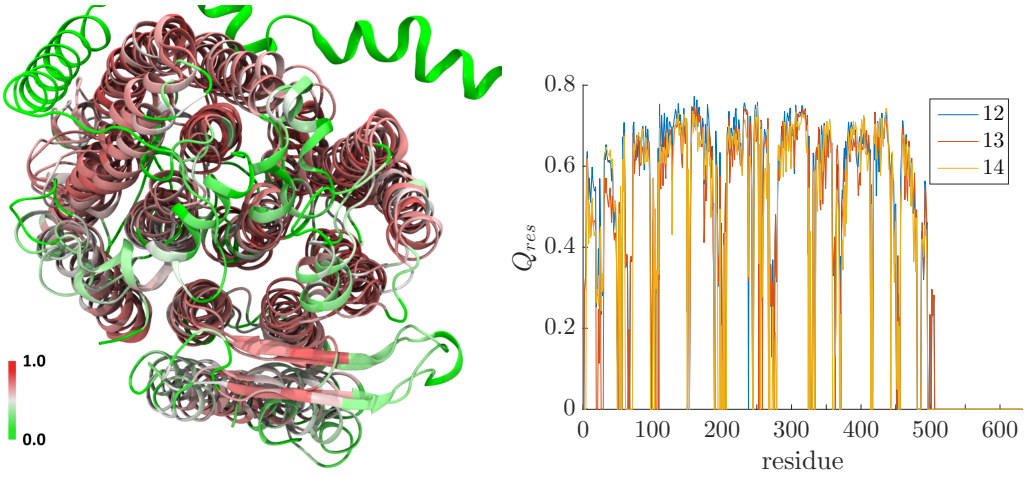
helix HL in a cylinder around helix HL. It estimates, if the percentage of water molecules above helix HL remains approximately constant through the simulation indicating the stability of the water dynamics around helix HL. For a suggested stabilising structure, consistency in behaviour is vital.

5. RESULTS AND DISCUSSION

This Chapter introduces the key results of this study. Structural and sequence analysis of the crystal structure is presented first, followed by stability analysis to ensure sufficient simulation length. Water dynamics analysis is presented for each subunit Nqo12-14, and the key residues are summarised. Discussion of the obtained results and their significance ends this Chapter.

5.1 Structural and sequence analysis

The membrane arm of complex I consists of 7 subunits, three of which are structurally homologous. In order to study the homology, and to determine which areas of the protein subunits are conserved, structural alignment was performed on the crystal structure of complex I from *T. thermophilus* (PBD id 4HEA [4]) for subunits Nqo12-14. Figure 5.1 shows the structural alignment for subunits Nqo12-14 viewed from the P-side. In Figure 5.1a the three subunits are shown superimposed on each other and coloured according to the Q_{res} value. Regions shown in red are structurally more similar than regions shown in green. Structural alignment indicates that the transmembrane helices are structurally similar and superimpose well, whereas the loops connecting the helices are less conserved structurally. The plot in Figure 5.1b shows Q_{res} against residue number in all three subunits. The value of Q_{res} alternates between 0 and approximately 0.6-0.7 with regular intervals. The appearance of the plot is linked to the structural similarity of helices in the subunits. Additionally, in the lower right corner of Figure 5.1a the connecting β-sheet structures, β hairpins, are well displayed and they as well have structural similarity between the subunits. The β-sheets have been suggested to play a role in stabilising the membrane domain and attaching the subunits to each other [8]. The residues 506 to 632 do not show structural alignment, because they correspond to helix HL, which is a unique feature of the subunit Nqo12.



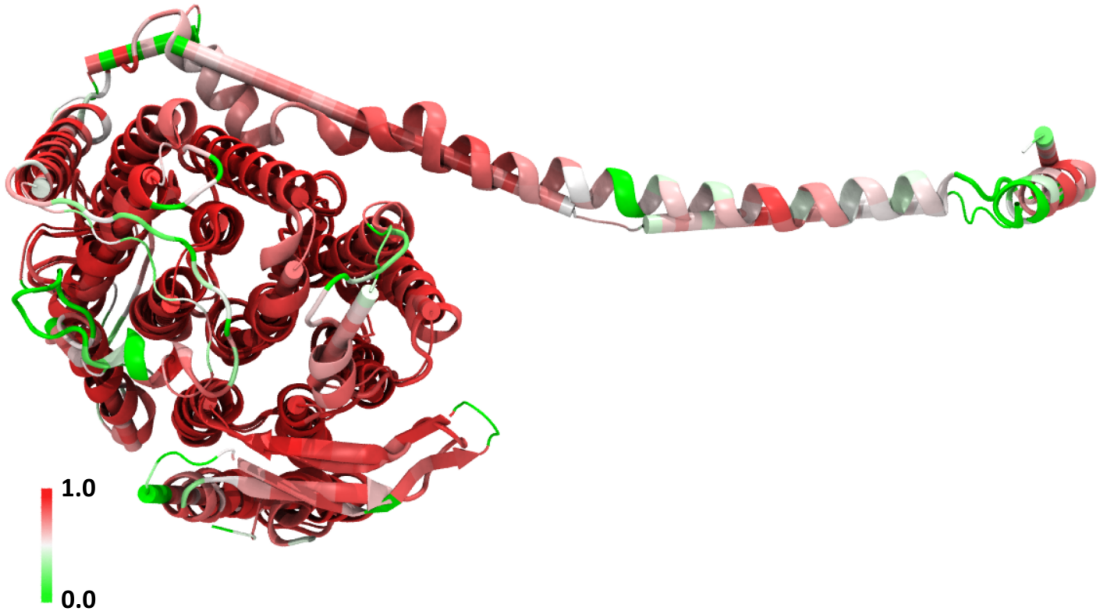
(a) Superimposed subunits Nqo12-14 coloured according to Q_{res} value.

(b) Q_{res} from structural alignment for each residue in subunits Nqo12-14.

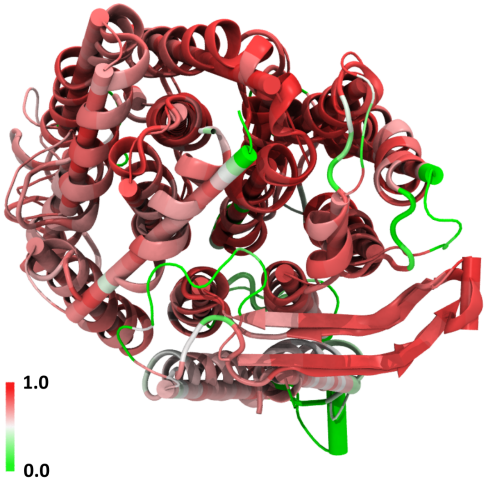
Figure 5.1 *Structural alignment. Figure 5.1a shows the superimposed structures of homologous subunits Nqo12-14 viewed from P-side coloured according to the Q_{res} value from green to red. Red regions are structurally similar and green are not. Figure 5.1b shows the value of Q_{res} for each residue.*

The two earlier reported computational studies related to the water dynamics of complex I were performed on the crystal structure resolved from *E. coli* [8, 53, 57] (PDB id 3RKO). Structure-based sequence alignment of the membrane domain of complex I has been done earlier [4, 8], and it shows a selection of well conserved charged residues. Structural alignment was carried out between crystal structures of complex I from *E. coli* and *T. thermophilus* (PDB ids 3RKO and 4HEA) in order to compare them. Structural alignment was performed for each subunit Nqo12-14 separately. The results are shown in Figure 5.2. The crystal structures superimpose very well. Additionally, careful inspection of the plots and images indicates that the most structurally conserved regions lie within the transmembrane helices, whereas the loops connecting the helices have structural variation.

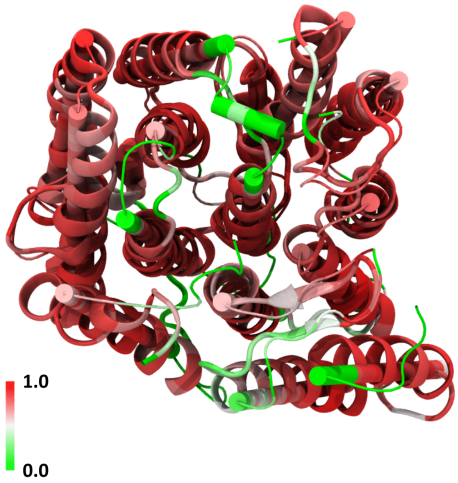
Recently, the crystal structure of mitochondrial complex I from *Yarrowia lipolytica* was reported [36]. The resolution of the structure is in the range of 3.6 to 3.9 Å, which is less than that of the crystal structure used in this study (3.3 Å) [4]. A number of amino acid residues are not resolved in the crystal structure of complex I from *Y. lipolytica*. Sequence data for complex I from *Y. lipolytica* is however available, which enables the generation of sequence alignment.



(a) Subunit Nqo12 coloured according to the value of Q_{res}.



(b) Subunit Nqo13 coloured according to the value of Q_{res}.



(c) Subunit Nqo14 coloured according to the value of Q_{res}.

Figure 5.2 *Structural alignment of E. coli and T. thermophilus complex I crystal structures coloured according to Q_{res} value. In each image, crystal structure from E. coli is shown in cartoon and T. Thermophilus in ribbons. The scale in the images is set so that green colour refers to structurally dissimilar areas and red colour to similar areas.*

In addition to the structural alignment, sequence alignment was also performed on sequence data of subunits Nqo12/ NuoL/ Nu5M, Nqo13/ NuoM/ Nu4M and Nqo14/ NuoN/ Nu2M from *T. thermophilus*, *E. coli*, and *Y. lipolytica*, respectively, to iden-

tify the conserved residues. The sequence data was retrieved from the UniProt database [112] [103, 113–120]. The sequence alignment is shown in Figures S.1, S.2 and S.3 (See Supplementary Material). The colouring scheme is such that the residues highlighted in pink are identical in the sequences. Secondary structure is assigned according to *T. thermophilus* structure as spiral on top the corresponding residues in Figures S.1, S.2 and S.3. From sequence alignment, it is observed that most of the conserved residues reside in the transmembrane helices. However, there are also non-conserved residues, which partly explain the obtained differences between the simulations and earlier studies (see more below).

Structurally conserved regions in complex I lie within the transmembrane helices both among the subunits as well as within the different crystal structures. The most conserved residues in complex I are concentrated to the middle region of the subunits. They constitute a wire of conserved residues through the membrane arm of the protein [8] suggesting a possible functional purpose. Structural and sequential similarities propose that the proton transfer pathways in subunits Nqo12-14 should resemble each other in complex I in different organisms.

Tables S.1, S.2 and S.3 (see Supplementary material) list the corresponding residues in subunits Nqo12/ NuoL/ Nu5M, Nqo13/ NuoM/ Nu4M and Nqo14/ NuoN/ Nu2M obtained from structure-based sequence alignment [8]. The residues in the table are gathered from three structural studies [4, 8, 36], all of which propose tentative proton transfer pathways. In these studies, the water pathways are suggested based on the conservation and chemical properties of the residues. The structural studies are carried out on enzymes from different organisms, and thus due to structural and sequential dissimilarities the suggested water pathways differ a bit. Residues suggested in each study are marked with italics. The MD simulation studies performed so far [53, 57] both suggest pathways similar to the structurally deducted pathways. These tables are helpful for comparison between this work and earlier studies.

5.2 Stability of the systems

How MD simulations function at a theoretical level is well understood, but in practise system construction and stability during simulations should be checked for each system. Therefore, the stability and convergence of the model systems must be assessed prior to further analysis. The model systems studied in this work are com-

prised of a protein and a membrane, and the stability of both of the components must be estimated.

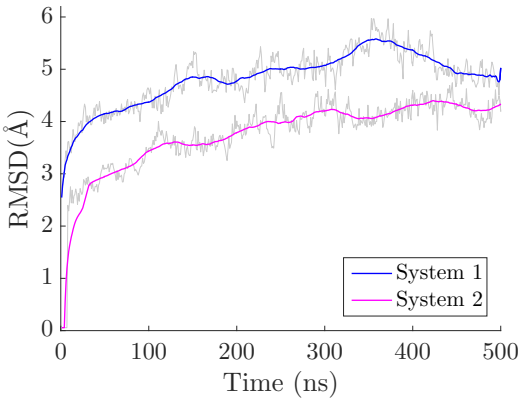
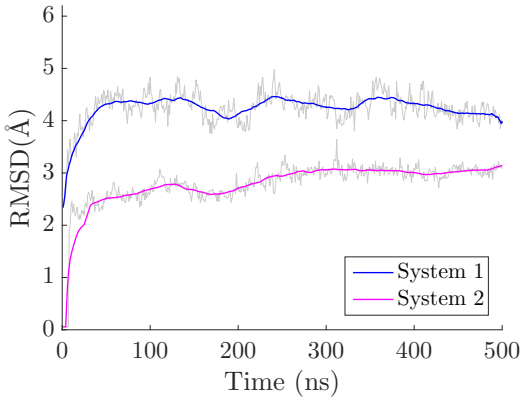


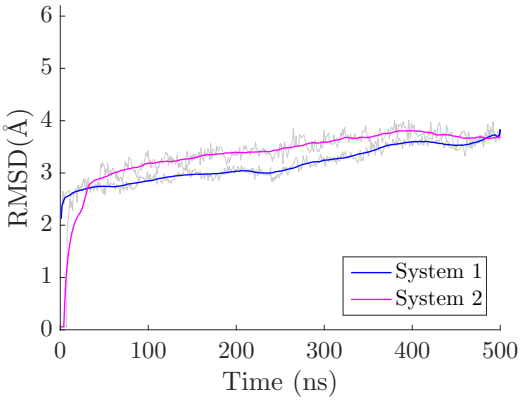
Figure 5.3 *RMSD of the Cα atoms of the entire complex I.*

Root mean square deviation (RMSD) was used for estimating the stability of the protein. Figure 5.3 shows the RMSD plot of Cα atoms computed from the aligned trajectories of model systems. The plot shows that RMSD values of the systems for the entire protein stabilise to approximately 5 $\AA$. However, RMSD of the membrane domain and hydrophilic domain was also plotted separately. The results are shown in Figure 5.4. These plots show that the membrane and the hydrophilic domain are stabilized. The RMSD plots do not show unrealistic conformational changes. Therefore, it can be concluded that the simulations are stable.

In order to analyse the stability of the membrane, the membrane thickness from the 500 ns MD simulations was calculated. The calculation was performed with the MEMBPLUGIN [121], and the results are shown in Figure 5.5a. In both of the model systems, the membrane thickness stabilises in the beginning of the MD simulation and then remains constant. For the first few nanoseconds, the membrane thickness is as high as 40 $\AA$ in both cases, but it stabilises to a value of approximately 38.1 $\AA$. The membrane thickness in both independent systems stabilises to nearly the same value, which indicates stability. The bilayer thickness of POPC membrane and its dependence on temperature have been determined experimentally [122]. The results are 39.1 ± 0.78 $\AA$ and 37.9 ± 0.8 $\AA$ at 30 and 50°C, respectively

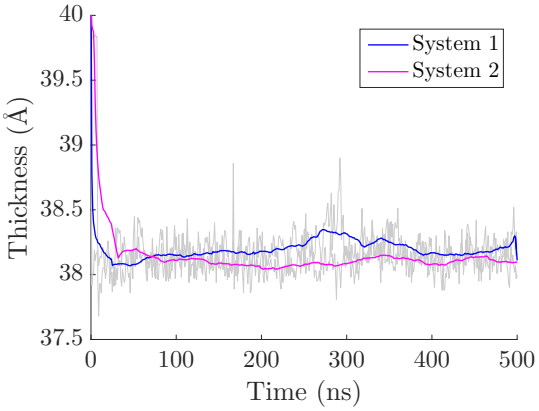


(a) RMSD of the Cα atoms of the membrane arm of complex I.

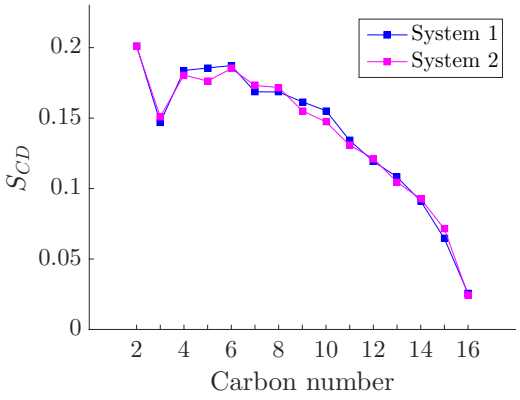


(b) RMSD of the Cα atoms of the hydrophilic arm of complex I.

Figure 5.4 *The plots show the RMSD value as a function of time for (a) the membrane domain of the protein and (b) the hydrophilic domain of the protein. The membrane domain includes subunits Nqo7, Nqo8 and Nqo10-14, and the hydrophilic domain subunits Nqo1-6, Nqo9, Nqo15 and Nqo16.*



(a) Membrane thickness.



(b) Deuterium order parameter.

Figure 5.5 *Membrane thickness and deuterium order parameter. Deuterium order parameter was measured for the palmitoyl chain of POPC.*

[122]. In this work, the temperature is 310 K (37°C) and membrane thickness 38.1 Å, so the model membrane is in reasonable agreement with experimental data.

In addition, area per lipids (APL) was estimated by approximating the protein structure in the membrane by a rectangle or a polygon. This analysis yields an approximate value of APL close to 70 $Å^2$. In earlier studies, APL values for pure POPC lipid bilayer were found to be 60 - 65 $Å^2$ [122–124]. The current results

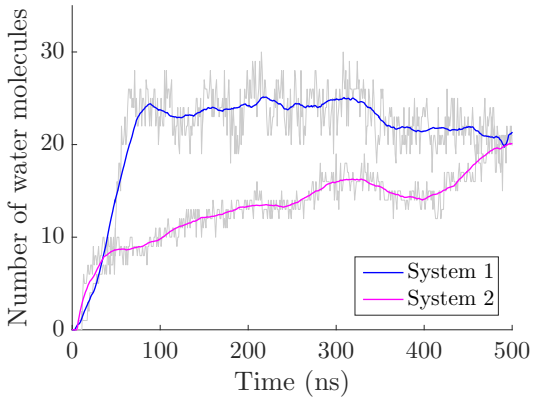
roughly match the available data, and slight variation is likely due to the unique protein structure.

In order to further validate the stability of the membrane, the deuterium order parameter (S_{CD}) for the palmitoyl chain of POPC (see Figure 5.5b) was measured. S_{CD} indicates how ordered the membrane is. The abscissa of Figure 5.5b refers to the carbon number in the palmitoyl chain in ascending order starting from the first CH_2-group. The carbon atoms closer to the head group of POPC are more ordered, than the ones in the chain. The headgroup atoms interact with each other and the surrounding water bulk, which makes motion of the carbon atoms close to the headgroup atoms restricted. The results agree well with experimental results [125], but there are small differences between the membrane used in this work and the earlier computational results [126]. The earlier computational results show slightly higher lipid order parameter than the results in Figure 5.5b. However, simulation lengths in earlier study were much shorter compared to this work, which may explain the difference. Additionally, the earlier study was performed with pure POPC membrane without embedded proteins. Nevertheless, the S_{CD} values of the membrane are in agreement with experimental data for liquid-phase POPC membrane. It can be concluded that the membrane is in liquid phase as expected physiologically.

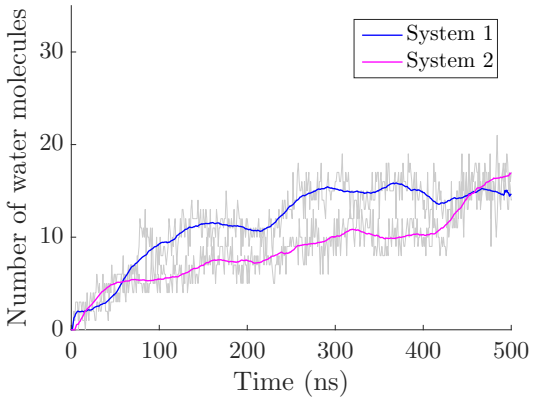
5.3 Water dynamics in the homologous subunits

In both model systems, a rapid and spontaneous hydration occurs in the membrane arm of complex I. The hydration pattern roughly follows a similar pattern in the independent simulation set-ups. Additionally, the formed water pathways are similar between the homologous subunits Nqo12-14.

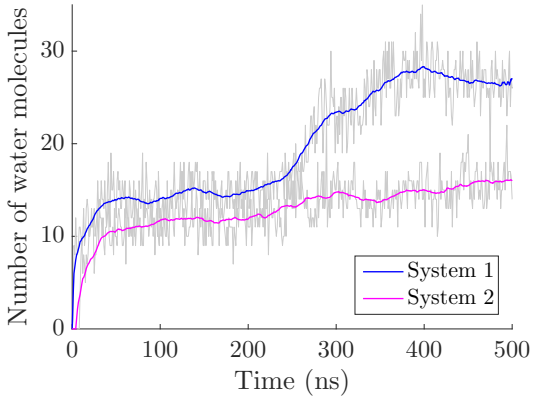
Figure 5.6 shows water occupancies in subunits Nqo12-14 in both the model systems. There are some differences in the number of internal water molecules between the two systems, but in Nqo12 and Nqo13, the water occupancies seem to stabilise as simulation length increases. Subunit Nqo14 on the contrary shows a difference in hydration towards the end of the simulations. Subunit Nqo14 is closest to the site for quinone binding, where the two model systems differ from each other, and thus it is possible that these differences are mechanistically relevant. Importantly, Figure 5.6 shows that it takes 10-100 ns for the water molecules to diffuse into the



(a) Subunit Nqo12.



(b) Subunit Nqo13.



(c) Subunit Nqo14.

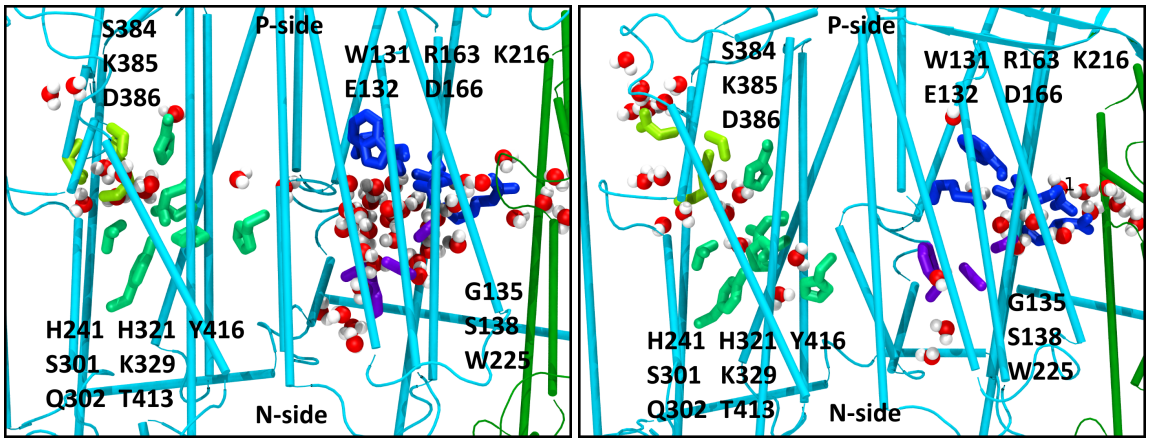
Figure 5.6 *Water occupancies in subunits Nqo12-14.*

subunits. Thus, most of the analysis performed in this thesis does not include the first 100 ns of the trajectories.

In the following subsections, I refer to connections to the P-side and the N-side as P-cluster and N-cluster, respectively.

5.3.1 Subunit Nqo12

In subunit Nqo12, water molecules are concentrated into two separate clusters, P-cluster and N-cluster. The clusters are located close to the charged residues in the middle plane of the membrane arm. The connection between the clusters is fairly transient, but in several frames of the simulation trajectories there is a water based connection between the two clusters, which enables the determination of the full



(a) Model system 1. **(b)** Model system 2.

Figure 5.7 *Water wire in subunit Nqo12. The helices are shown as cartoon with cyan in subunit Nqo12 and green in subunit Nqo13. Water molecules are shown in red-white spheres. The sidechains of residues involved in the water channel coordination are shown separately. Residues connecting to the N-side are shown in purple, to the adjacent subunit in dark blue, to the P-side in yellow, and the ones in the middle in green.*

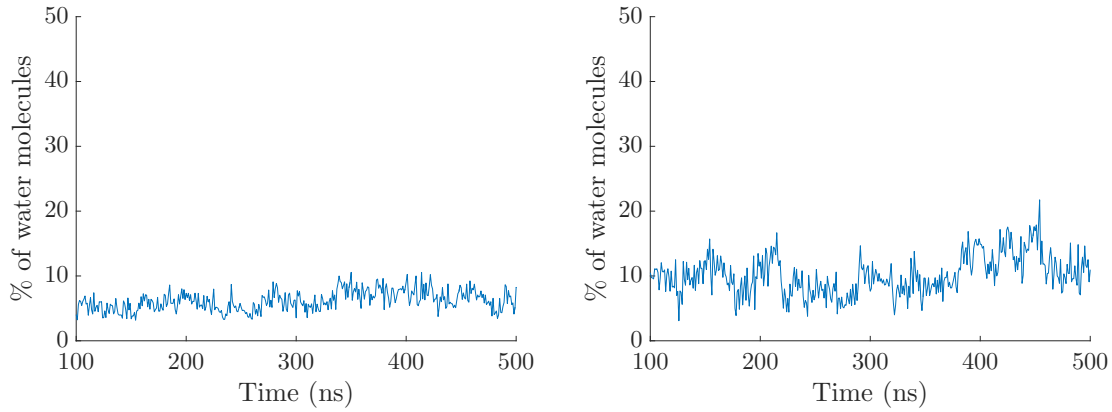
water pathway through the subunit (see Figure 5.7a).

Water pathway through subunit Nqo12 is S-shaped. In order to provide quantitative proof of the water wires and locate the exact residues involved in forming them, hydrogen bond analysis was performed. The internal water molecules in subunit Nqo12 were chosen and the hydrogen bonding between them and the protein was measured. The results are shown in Table S.4 (see Supplementary Material). The results of hydrogen bond analysis for subunit Nqo12 are very similar between the two model systems. From the table, the residues that very likely participate in water channel formation are identified. However, this table only lists residues that are hydrogen bonded to the internal water molecules for more than 10 % of the simulation time due to the high number of hydrogen bonds during the course of the simulation. Thus, in addition to choosing residues from the table, the trajectory was visually analysed and searched for residues that border the water molecules inside subunit Nqo12. It is to note that residues that are strongly hydrophobic or not highly conserved were excluded.

Residues Gly135, Ser138 and Trp225 form the connection to N-side. The connection is in close proximity to the bulk water, where water molecules move freely, and thus no stable hydrogen bonds are formed. Residues Trp131, Glu132, Arg163, Asp166

and Lys216 are somewhat between the middle of the subunit and the residues close to the N-side. Together they form a connection from the bulk water to the middle of the membrane arm. Residues Glu132, Arg163, Asp166 and Lys216 constitute a cluster of conserved charged residues in the middle plane of the membrane forming a small hydrophilic site inside the subunit. The high conservation of this region indicates a possible functional role. In fact, it was observed that water molecules flow to the adjacent subunit Nqo13 through this region in conjunction with the highly conserved residue Glu377 (Nqo13). The connection between subunits is strong, and may play an important role for instance in the coupling mechanism. Residues His241, Ser301, Gln302, His321, Lys329, Thr413 and Tyr416, which were obtained from the hydrogen bond analysis, lie in the middle plane of the subunit forming a chain of conserved residues. They connect the N-cluster with the P-cluster, which is bordered by residues Ser384, Lys385,and Asp386. The residues identified in this work form a chain and together with water molecules allow identification of a proton transfer pathway. The observed hydration pattern agrees well with the earlier studies.

Point mutation E144Q in NuoL (E132 in Nqo12) reduces the proton pumping activity to 10 % of the wild type [110]. The mutation also reduces the oxidoreductase activities to 15 % of the wild type [110]. The side chains of glutamic acid and glutamine are chemically different. Charged residues attract water strongly, hence this point mutation is likely to decrease hydration and weaken hydrogen bonding in subunit Nqo12. More transient hydrogen bond network may prevent the efficient proton transfer, hence decreased the activity of complex I. Another remarkable oxidoreductase activity defect is related to mutation K229R in NuoL (K216 in Nqo12), which reduces the activity to 30 % of the wild type, respectively [110]. It was obtained that residues Glu132 and Lys216 are part of a small hydrophilic region that allows water molecules to flow both through subunit Nqo12 as well as to the adjacent subunit Nqo13. The defects caused by the point mutations are significant to the function of the enzyme; hence, they indicate that Glu132 and Lys216 have a functional role. Earlier studies have also proposed these residues to be part of the key residues in proton pumping, and the simulations support this view. Mutation D400A/N/E in NuoL (D386 in Nqo12) decreases oxidoreductase activities to 70 % /90 % /100 % of the wild type and proton pumping activity to 50 % /70 % /90 % [110]. Because alanine is a hydrophobic amino acid, it evidently has a stronger effect on proton pumping than glutamic acid, whose side chain is charged like aspartic acid. As-



(a) Model system 1. **(b)** Model system 2.

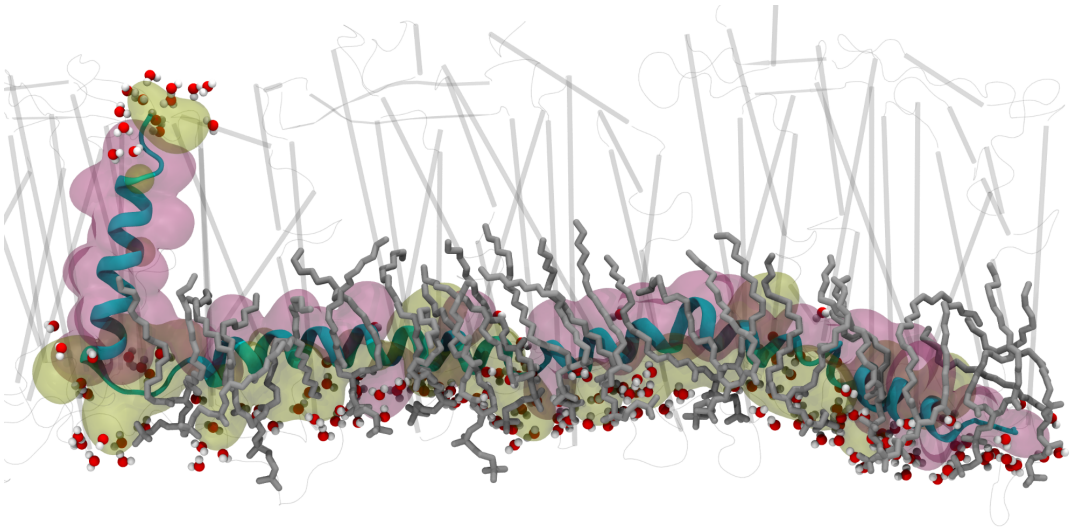
Figure 5.8 *Hydration around helix HL. The plots show the percentage of water molecules above helix HL in both model systems.*

paragine is hydrophilic in nature, hence it does not alter the proton pumping rate as strongly as the hydrophobic alanine. However, even with alanine in place of aspartic acid, the enzyme does not lose its functionality completely. The efficiency is severely affected, but complex I stays somewhat functional, which supports the view of a chain of multiple residues participating in proton pumping in complex I. It can be concluded that a single point mutation will not cease the energy production.

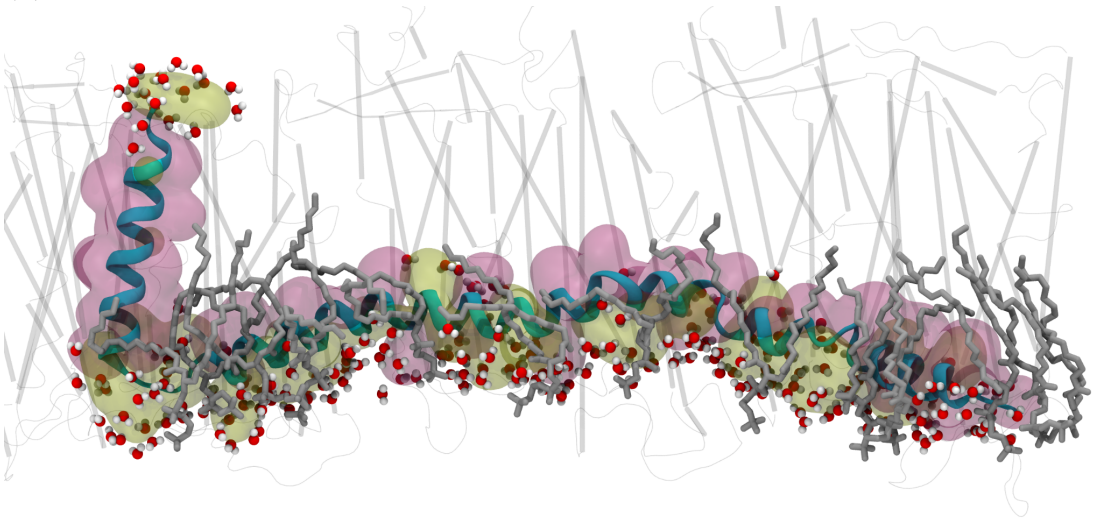
Helix HL is a long horizontal helix, which is unique to subunit Nqo12. Based on its structure and location, it has been suggested to play a role in stabilising the protein structure and keeping subunit Nqo12-14 attached to each other [4, 8]. It is not fully understood, how the stabilisation would take place. Earlier studies also propose that helix HL acts as a coupling element [8, 127]. However, it has been shown that helix HL does not play a role in proton pumping [128].

It is observed that helix HL borders the bulk water and the membrane. It interacts with the water molecules on one side, whereas with the acyl chains of the membrane lipids on the other. Figure 5.9 shows the location and surroundings of this helix in both the model systems. In order to estimate how stable the water dynamics around helix HL are, hydration analysis according to Section 4.2.4 was performed. Figure 5.8 shows the results of analysis described in Section 4.2.4. In both model systems, it is observed that the hydration is approximately 10 % and remains constant through the simulation. The plots show that the pattern remains similar through the simulations, and the hydration along helix HL is consistent through the simu-

lations. Based on these simulations, it is suggested that the helix prevents excess water molecules from entering the space between subunits. Excess water molecules in critical locations may weaken the non-bonded interactions in between the subunits, thereby leading to destabilization of protein structure. Additionally, helix HL interacts with the subunits through polar interactions, hence physically stabilising the subunits.



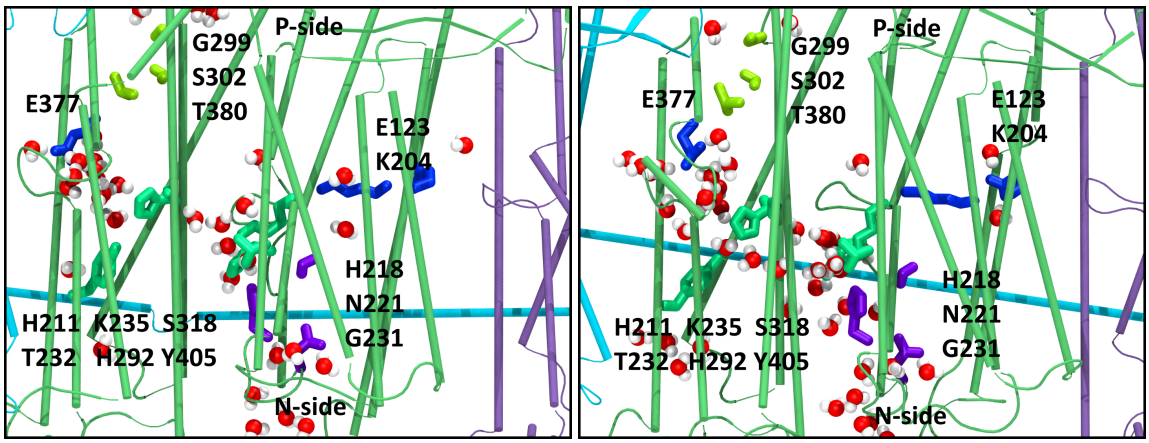
(a) Model system 1.



(b) Model system 2.

Figure 5.9 *Hydration and interaction around helix HL. The helices of subunits Nqo12-14 are shown in the background with transparent grey. The backbone of helix HL is shown with cyan ribbons. The hydrophilic side chains of helix HL are shown in yellow surface, and the hydrophobic ones in purple surface. Water molecules surrounding helix HL are shown as red-white spheres, and the lipids surrounding helix HL in grey. The hydrophilic side chains of helix HL interact with water molecules and the head groups of the lipid membrane, while the hydrophobic side chains interact with the acyl chains of the lipids.*

5.3.2 Subunit Nqo13



(a) Model system 1. **(b)** Model system 2.

Figure 5.10 *Water wire in subunit Nqo13. The helices are shown as cartoon with green in subunit Nqo13, cyan in subunit Nqo12, and lilac in subunit Nqo14. Water molecules are shown in red-white spheres. The side chains of residues involved in the water channel coordination are shown separately. Residues that connect to the N-side are shown in purple, to the adjacent subunits in dark blue, to the P-side in yellow, and the ones in the middle in green.*

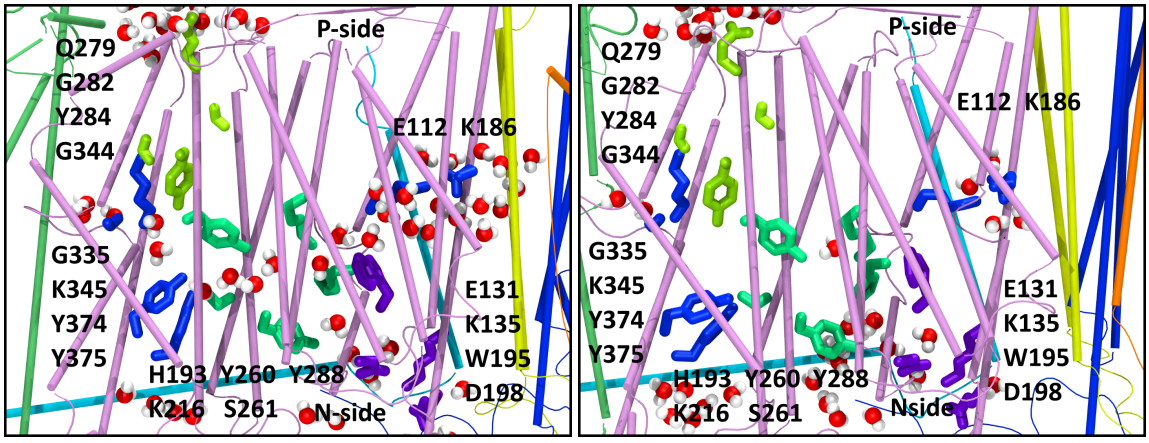
The water molecules that diffuse into subunit Nqo13 from Nqo12 perturb the water wire formation in Nqo13, and the connections to bulk water in the N-side and the P-side are transient and not as clear as in Nqo12. The middle region of Nqo13 is however well hydrated, partly due to the flow of water molecules from subunit Nqo12. Highly conserved residue Glu377 guides the water molecules from the middle of subunit Nqo12 towards the middle of subunit Nqo13. Residues His211, Thr232, Lys235, His292, Ser318 and Tyr405 form a continuous connection through the middle plane of subunit Nqo13 from Nqo12 to Nqo14 (see Figure 5.10). Residues Glu112 and Lys204 complete the connection and allow transfer of water molecules towards subunit Nqo14. The transient connection to the N-side is coordinated by residues His218, Asn221, and Gly231, and to the P-side by Gly299, Ser302, and Thr380.

Mutation data of NuoM indicate that the mutations E144A/Q in NuoM (E123 in Nqo13) lead to a drastic decrease in oxidoreductase activities, and no proton pumping [106, 107, 109]. On the other hand, mutation E144D has no effect on proton pumping. Glutamic acid and aspartic acid differ in side chain length, whereas alanine is remarkably different. Alanine will generate a much more transient hydrogen bond network in comparison to glutamic acid or aspartic acid, which will

decrease the proton transfer rate. Additionally, mutation K234A in NuoL (K204 in Nqo13) results in significantly decreased proton pumping and 5-10 % oxidoreductace activities. Point mutations H241A/E/K/R (His211) also lead to a decrease in oxidoreductase activities and proton pumping [106, 111]. Point mutation K265A (Lys235), too, has a detrimental effect on proton pumping [106, 107]. Lysine and histidine are a hydrophilic amino acids, whereas alanine is hydrophobic, which explains the experimental results.

Hydrophilic residues attract water molecules and stabilise the proton transfer pathways for longer time scales. Mutations, upon which hydrophobic residues replace hydrophilic ones, lead to a severe loss of enzymatic efficiency. Analysis of the simulations in this work on wild-type enzyme, and stability analysis of proton transfer pathways, explains the observed loss of activity upon mutations, which is most likely due to perturbation in the H-bond network (proton transfer).

5.3.3 Subunit Nqo14



(a) Model system 1. **(b)** Model system 2.

Figure 5.11 *Water wire in subunit Nqo14. The helices are shown as cartoon with lilac in subunit Nqo13, green in subunit Nqo13, and orange, yellow and blue in subunits Nqo7, Nqo10 and Nqo11, respectively. Water molecules are shown in red-white spheres. The sidechains of residues involved in the water channel coordination are shown separately. Residues that connect to the N-side are shown in purple, to the adjacent subunits in dark blue, to the P-side in yellow, and the ones in the middle in green.*

Water molecules distribute throughout the subunit Nqo14 starting from the hydration in subunits Nqo7, Nqo10 and Nqo11 and towards subunit Nqo13. Similarly to subunit Nqo13, subunit Nqo14 also does not have a steady water based connection

through the subunit from the N-side to the P-side. Rather, there is a water based connection in the middle plane of the subunit that does not reach the water bulk on the N-side or the P-side. Residues Glu112 and Lys186 form the connection to three small adjacent subunits Nqo7, Nqo10 and Nqo11 in the middle plane of the subunit. Residues Gly335, Lys345, Tyr374 and Tyr375 form a similar connection towards the subunit Nqo13. Two water molecule clusters are formed around these residues, the N-cluster and the P-cluster, respectively. The two clusters are connected via a selection of residues in the middle of Nqo14: His193, Lys216, Tyr260, Ser261 and Tyr288. The transient connection to the N-side is coordinated by residues Glu131, Lys135, Trp195, and Asp198, and to P-side by residues Gln279, Gly282, Tyr284, and Gly344.

Interestingly, mutation data of subunit NuoN indicate that the point mutations E133A/C/D in NuoN (Glu112 in Nqo14) have no effect on proton pumping and only a slight decrease in oxidoreductase activities [103]. Mutation K217R in NuoN (Lys186 in Nqo14) also causes only a slight decrease in proton pumping activity. Both lysine and arginine are basic amino acids; hence, the effect of mutation is expected. Point mutation K247C in NuoN (Lys216 in Nqo14) decreases proton pumping activity to half of the wild type, while K247R has no major effect [103], due to the chemical properties of the side chains. Additionally, point mutations K395C/R in NuoN (Lys345 in Nqo14) drop the oxidoreductase activities to 5 % and 30 % of the original value [103]. Overall, mutation data of subunit NuoN is in agreement with the observed results.

5.3.4 Outline

The tentative pathways for proton pumping obtained from the simulations are summarised in Table 5.1.

Table 5.1 *Residues involved in the water wires suggested based on the simulations. See main text for each subunit for additional information.*

Subunit	Details	Residues
Nqo12	Connection to N-side	Gly135, Ser138, Trp225
	Connection to subunit Nqo13	Trp131, Glu132, Arg163, Asp166, Lys216
	Middle residues	His241, Ser301, Gln302, His321, Lys329, Thr413, Tyr416
	Connection to P-side	Ser384, Lys385, Asp386
Nqo13	Connection to N-side	His218, Asn221, Gly231
	Connection to subunit Nqo14	Glu123, Lys204
	Middle residues	His211, Thr232, Lys235, His292, Ser318, Tyr405
	Connection to subunit Nqo12	Glu377
	Connection to P-side	Gly299, Ser302, Thr380
Nqo14	Connection to N-side	Glu131, Lys135, Trp195, Asp198
	Connection to subunits Nqo7/10/11	Glu112, Lys186
	Middle residues	His193, Lys216, Tyr260, Ser261, Tyr288
	Connection to subunit Nqo13	Gly335, Lys345, Tyr374, Tyr375
	Connection to P-side	Gln279, Gly282, Tyr284, Gly344

5.4 Discussion

The mechanism by which electron transfer is coupled to proton pumping in complex I has remained elusive. There are also many other open questions concerning complex I research, for example quinone binding sites and the presence of the fourth channel. Here the focus is on water dynamics in the antiporter-like subunits of complex I using atomistic molecular dynamics simulations.

Earlier studies regarding water dynamics in respiratory complex I were performed on a partial structure of complex I core including only six of the membrane subunits. Here, water dynamics in the antiporter-like subunits are studied from molecular dynamics simulations performed on the entire structure of respiratory complex I, including all of the core subunits, and two additional organism specific subunits. The observed results agree with the earlier results. Additionally, the results were cross-checked with mutagenesis data to verify that the computational models agree with experimental results. However, water molecules do not disperse freely through the structure during the simulations. Most interestingly, a strong water based connection is obtained through the middle plane of the membrane arm. The connection reaches all the way through the antiporter-like subunits in both model systems suggesting a possible functional role. Earlier structural studies of complex I have shown a line of conserved titrable residues in the middle plane of the membrane arm. The obtained water based connection runs through the line of titrable residues leading to the proposal that these residues along with the water molecules may serve as a possible coupling route through the subunits. Overall, mutation data of the membrane domain shows that mutating the titrable residues to non-titrable residues will have a remarkable effect on either oxidoreductase activities, proton pumping activity or both. However, it is to be noted that the enzyme will not become completely dysfunctional with a single point mutation, which is in agreement with the simulation data that a large number of residues and water molecules participate in proton channel formation.

The study of water dynamics was extended to the unique horizontal helical structure in subunit Nqo12. Helix HL is an amphipathic long helix, which has been suggested to play a role in stabilising the membrane arm of complex I. The hydrophilic side of the HL interacts with water molecules, while the hydrophobic side interacts with the acyl chain tails of the membrane lipids. It is observed that the hydration pattern around the helix remains stable through both of the simulations. The results suggest

that helix HL stabilises the membrane arm by regulating the water molecule flow to the space between the protein and membrane.

6. CONCLUSIONS

The function of the respiratory complex I is essential in energy production of cells. Complex I is an oxidoreductase, which couples electron transfer to proton pumping via a mechanism that has remained elusive. Dysfunction of complex I and mutations in its structure are known to cause severe diseases and play an important role in, for example, ageing. So far the proton pumping mechanism of complex I has been unclear and only two earlier MD studies related to it have been performed on a partial crystal structure of complex I.

The aim of this thesis was to perform MD simulations on the structure of the entire complex I and to study the water dynamics of the three homologous membrane subunits. Thorough comparison between the so far published structural and MD studies was done and the results were compared. A set of residues participating in water wire formation were observed for each antiporter-like subunit. Additionally, a water-mediated connection accompanied by a selection of protonable residues through the middle plane of the membrane was observed. This result elucidates the possible coupling route in the membrane arm subunits. Additionally, water dynamics along helix HL in subunit Nqo12 propose a possible stabilization mechanism.

The analysis performed in this thesis will be continued to study the fourth possible proton pumping pathway, which lies closer to the interface of the membrane and hydrophilic arm. Computational methods also provide possibilities to analyse the behaviour and dynamics of quinone as well as for instance tentative coupling mechanisms. The most recent atomic structures of mammalian complex I provide a possibility to study the role of the supernumerary subunits computationally.

Complex I is a massive, fascinating molecular machine. Its mechanism and function will provide interesting research topics for several years to come.

BIBLIOGRAPHY

[1] **Campbell, M.K. and Farrell, S.O.**, *Biochemistry*. Brooks/Cole cengage learning, 7th edition. 2012.

[2] **Alberts, B., Bray, D., Hopkin, K., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P.**, *Essential cell biology*. Garland Science, 3rd edition. 2010.

[3] **Sazanov, L.A. and Hinchliffe, P.**, Structure of the Hydrophilic Domain of Respiratory Complex I from *Thermus thermophilus*. *Science*, 311: 1430–1436. 2006.

[4] **Baradaran, R., Berrisford, J.M., Minhas, G.S., and Sazanov, L.A.**, Crystal structure of the entire respiratory complex I. *Nature*, 494: 443–448. 2013.

[5] **Nicholls, D.G. and Ferguson, S.J.**, *Bioenergetics*. Academic Press, 4th edition. 2013.

[6] **Nelson, P.**, *Biological physics: energy, information, life*. W. H. Freeman and Company, updated 1st edition. 2008.

[7] **Branden, C. and Tooze, J.**, *Introduction to protein structure*. Garland Publishing, 2nd edition. 1999.

[8] **Efremov, R.G. and Sazanov, L.A.**, Structure of the membrane domain of respiratory complex I. *Nature*, 476: 414–420. 2011.

[9] **Rogers, B., Pennathur, S., and Adams, J.**, *Nanotechnology: understanding small systems*. CRC Press, 2nd edition. 2011.

[10] **Barnes, G.T. and Gentle, I.R.**, *Interfacial science: an introduction*. Oxford University Press, 2nd edition. 2011.

[11] **Madigan, M.T., Martinko, J.M., Dunlap, P.V., and Clark, D.P.**, *Brock Biology of microorganisms*. Pearson Benjamin Cummings, 12th edition. 2009.

[12] **Lane, N. and Martin, W.**, The energetics of genome complexity. *Nature*, 467: 929–934. 2010.

[13] **Kühlbrandt, W.**, Structure and function of mitochondrial membrane protein complexes. *BMC Biology*. 2015.

[14] **Welte, W., Nestel, U., Wacker, T., and Diederichs, K.**, Structure and function of the porin channel. *Kidney International*, 48: 930–940. 1995.

[15] **Llopis, J., McCaffery, J.M., Miyawaki, A., Farquhar, M.G., and Tsien, R.Y.**, Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 95: 6803–6808. 1998.

[16] **Srere, P.A.**, The infrastructure of the mitochondrial matrix. *Trends in Biochemical Sciences*, 5: 120–121. 1980.

[17] **Partikian, A., Ölveczky, B., Swaminathan, R., Li, Y., and Verkman, A.**, Rapid diffusion of green fluorescent protein in the mitochondrial matrix. *Journal of Cell Biology*, 140: 821–829. 1998.

[18] **McKinney, E.A. and Oliveira, M.T.**, Replicating animal mitochondrial DNA. *Genetics and Molecular Biology*, 36: 308–315. 2013.

[19] **Gaston, D., Tsaousis, A., and Roger, A.**, Predicting proteomes of mitochondria and related organelles from genomic and expressed sequence tag data. *Methods in Enzymology*, 457: 21–47. 2009.

[20] **Mootha, V., Bunkenborg, J., Olsen, J., Hjerrild, M., Wisniewski, J., Stahl, E., Bolouri, M., Ray, H., Sihag, S., Kamal, M., Patterson, N., Lander, E., and Mann, M.**, Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. *Cell*, 115: 629–640. 2003.

[21] **Pagliarini, D., Calvo, S., Chang, B., Sheth, S., Vafai, S., Ong, S., Walford, G., Sugiana, C., Boneh, A., Chen, W., Hill, D., Vidal, M., Evans, J., Thorburn, D., Carr, S., and Mootha, V.**, A mitochondrial protein compendium elucidates complex I disease biology. *Cell*, 134: 112–123. 2008.

[22] **Sickmann, A., Reinders, J., Wagner, Y., Joppich, C., Zahedi, R., Meyer, H., Schönfisch, B., Perschil, I., Chacinska, A., Guiard, B., Rehling, P., Pfanner, N., and Meisinger, C.**, The proteome of *Saccharomyces cerevisiae* mitochondria. *Proceedings of the National Academy of Sciences of the United States of America*, 100: 13207–13212. 2003.

[23] **Nunnari, J. and Suomalainen, A.**, Mitochondria: in sickness and in health. *Cell*, 148: 1145–1159. 2012.

[24] **Neupert, W. and Herrmann, J.**, Translocation of proteins into mitochondria. *Annual Review of Biochemistry*, 76: 723–749. 2007.

[25] **Sharma, V.**, *Computational investigation of oxygen reduction and proton pumping in cbb_3-type Cytochrome c Oxidases*. Academic dissertation, University of Helsinki. 2012.

[26] **Pereira, M.M., Santana, M., and Teixeira, M.**, A novel scenario for the evolution of haem-copper oxygen reductases. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1505: 185–208. 2001.

[27] **Dudkina, N.V., Sunderhaus, S., Braun, H.P., and Boekema, E.J.**, Characterization of dimeric ATP synthase and cristae membrane ultrastructure from *Saccharomyces* and *Polytomella* mitochondria. *FEBS Letters*, 580: 3427–3432. 2006.

[28] **Schägger, H. and Pfeiffer, K.**, Supercomplexes in the respiratory chains of yeast and mammalian mitochondria. *EMBO Journal*, 19: 1777–1783. 2000.

[29] **Eubel, H., Heinemayer, J., Sunderhaus, S., and Braun, H.P.**, Respiratory chain supercomplexes in plant mitochondria. *Plant Physiology and Biochemistry*, 42: 937–942. 2004.

[30] **Wittig, I., Carrozzo, R., Santorelli, F.M., and Schägger, H.**, Supercomplexes and subcomplexes of mitochondrial oxidative phosphorylation. *Biochimica et Biophysica Acta*, 2006: 1066–1072. 2006.

[31] **Lenaz, G. and Genova, M.L.**, Structure and organization of mitochondrial respiratory complexes: a new understanding of an old subject. *Antioxidants & Redox Signaling*, 12: 961–1008. 2010.

[32] **Bultema, J.B., Braun, H.P., Boekema, E.J., and Kouril, R.**, Megacomplex organization of the oxidative phosphorylation system by structural analysis of respiratory supercomplexes from potato. *Biochimica et Biophysica Acta*, 1787: 60–67. 2009.

[33] **Genova, M.L. and Lenaz, G.**, Functional role of mitochondrial respiratory supercomplexes. *Biochimica et Biophysica Acta*, 1837: 427–443. 2014.

[34] **Sazanov, L.**, ed., *A structural perspective on respiratory complex I: structure and function of NADH:ubiquinone oxidoreductase*, chapter Structure of complex I, pp. 3–21. Springer. 2012.

[35] **Vinothkumar, K.R., Zhu, J., and Hirst, J.**, Architecture of mammalian respiratory complex I. *Nature*, 515: 80–84. 2014.

[36] **Zickermann, V., Wirth, C., Nasiri, H., Siegmund, K., Schwalbe, H., Hunte, C., and Brandt, U.**, Mechanistic insight from the crystal structure of mitochondrial complex I. *Science*, 347: 44–49. 2015.

[37] **Zhu, J., Vinothkumar, K.R., and Hirst, J.**, Structure of mammalian respiratory complex I. *Nature*, 536: 354–358. 2016.

[38] **Fiedorczuk, K., Letts, J.A., Degliesposti, G., Kaszuba, K., Skehel, M., and Sazanov, L.A.**, Atomic structure of the entire mammalian mitochondrial complex I. *Nature*. 2016, http://dx.doi.org/10.1038/nature19794, In press.

[39] **Yip, C., Harbour, M.E., Jayawardena, K., Fearnley, I.M., and Sazanov, L.A.**, Evolution of respiratory complex I: "supernumerary" subunits are present in the α-proteobacterial enzyme. *The Journal of Biological Chemistry*, 286: 5023–5033. 2011.

[40] **Carroll, J., Fearnley, I.M., Skehel, J.M., Shannon, R.J., Hirst, J., and Walker, J.E.**, Bovine complex I Is a complex of 45 different subunits. *The Journal of Biological Chemistry*, 281: 32724–32727. 2006.

[41] **Berrisford, J.M. and Sazanov, L.A.**, Structural basis for the mechanism of respiratory complex I. *Journal of Biological Chemistry*, 284: 29773–29783. 2009.

[42] **Page, C.C., Moser, C.C., Chen, X., and Dutton, P.L.**, Natural engineering principles of electron tunnelling in biological oxidation-reduction. *Nature*, 402: 47–52. 1999.

[43] **Verkhovskaya, M.L., Belevich, N., Euro, L., Wikström, M., and Verkhovsky, M.I.**, Real-time electron transfer in respiratory complex I. *Proceedings of the National Academy of Sciences of the United States of America*, 105: 3763–3767. 2008.

[44] **Kozlíková, B., Šebestová, E., Šustr, V., Brezovský, J., Strnad, O., Daniel, L., Bednář, D., Pavelka, A., Maňák, M., Bezděka, M., Beneš, P., Kotry, M., Gora, A., Damborský, J., and Sochor, J.**, CAVER Analyst 1.0: Graphic tool for interactive visualization and analysis of tunnels and channels in protein structures. *Bioinformatics*, 30: 2684–2685. 2014.

[45] **Pavelka, A., Šebestová, E., Kozlíková, B., Brezovský, J., Sochor, J., and Damborský, J.**, CAVER: algorithms for analyzing dynamics of tunnels in macromolecules. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 13: 505–517. 2016.

[46] **Sharma, V., Belevich, G., Gamiz-Hernandez, A.P., Róg, T., Vattulainen, I., Verkhovskaya, M.L., Wikström, M., Hummer, G., and Kaila, V.R.I.**, Redox-induced activation of the proton pump in the respiratory complex I. *Proceedings of the National Academy of Sciences of the United States of America*, 112: 11571–11576. 2015.

[47] **Kervinen, M., Pätsi, J., Finel, M., and Hassinen, I.E.**, A pair of membrane-embedded acidic residues in the NuoK subunit of *Escherichia coli* NDH-1, a counterpart of the ND4L subunit of the mitochondrial complex I, are required for high ubiquinone reductase activity. *Biochemistry*, 43: 773–781. 2004.

[48] **Kao, M.C., Nakamaru-Ogiso, E., Matsuno-Yagi, A., and Yagi, T.**, Characterization of the membrane domain subunit NuoK (ND4L) of the NADH-Quinone oxidoreductase from *Escherichia coli*. *Biochemistry*, 44: 9545–9554. 2005.

[49] **Mathiesen, C. and Hägerhäll, C.**, Transmembrane topology of the NuoL, M and N subunits of NADH:quinone oxidoreductase and their homologues among membrane-bound hydrogenases and bona fide antiporters. *Biochimica et Biophysica Acta*, 1556: 121–132. 2002.

[50] **Birrell, J.A. and Hirst, J.**, Truncation of subunit ND2 disrupts the threefold symmetry of the antiporter-like subunits in complex I from higher metazoans. *FEBS Letters*, 584: 4247–4252. 2010.

[51] **Vinothkumar, K.R. and Henderson, R.**, Structures of membrane proteins. *Quarterly Reviews of Biophysics*, 43: 65–158. 2010.

[52] **Knuuti, J., Belevich, G., Sharma, V., Bloch, D.A., and Verkhovskaya, M.**, A single amino acid residue controls ROS production in the respiratory Complex I from *Escherichia coli*. *Molecular Microbiology*, 90: 1190–1200. 2013.

[53] **Kaila, V.R.I., Wikström, M., and Hummer, G.**, Electrostatics, hydration, and proton transfer dynamics in the membrane domain of respiratory complex I. *Proceedings of the National Academy of Sciences of the United States of America*, 111: 6988–6993. 2014.

[54] **Agmon, N.**, The Grotthuss mechanism. *Chemical Physics Letters*, 244: 456–462. 1995.

[55] **Cukierman, S.**, Et tu, Grotthuss! and other unfinished stories. *Biochimica et Biophysica Acta*, 1757: 876–885. 2006.

[56] **Zumdahl, S.S. and DeCoste, D.J.**, *Chemical principles*. Brooks/Cole cengage learning, 7th edition. 2013.

[57] **Tan, P., Feng, Z., Zhang, L., Hou, T., and Li, Y.**, The mechanism of proton translocation in respiratory complex I from molecular dynamics. *Journal of Receptors and Signal Transduction*, 35: 170–179. 2014.

[58] **Murphy, M.P.**, How mitochondria produce reactive oxygen species. *Biochemical Journal*, 417: 1–13. 2009.

[59] **Dawson, T. and Dawson, V.L.**, Molecular pathways of neurodegeneration in Parkinson's disease. *Science*, 302: 819–822. 2003.

[60] **Balaban, R., Nemoto, S., and Finkel, T.**, Mitochondria, oxidants, and aging. *Cell*, 120: 483–495. 2005.

[61] **Sazanov, L.A.**, Respiratory complex I: mechanistic and structural insights provided by the crystal structure of the hydrophilic domain. *Biochemistry*, 46: 2275–2288. 2007.

[62] **Schapira, A.**, Human complex I defects in neurodegenerative diseases. *Biochimica et Biophysica Acta*, 1364: 261–270. 1998.

[63] **Wang, X., Wang, W., Li, L., Perry, G., gon Lee, H., and Zhu, X.**, Oxidative stress and mitochondrial dysfunction in Alzheimer's disease. *Biochimica et Biophysica Acta*, 1842: 1240–1247. 2014.

[64] **Yan, M.H., Wang, X., and Zhu, X.**, Mitochondrial defects and oxidative stress in Alzheimer disease and Parkinson disease. *Free Radical Biology & Medicine*, 62: 90–101. 2013.

[65] **Karplus, M. and McCammon, J.A.**, Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, 9: 646–652. 2002.

[66] **Paquet, E. and Viktor, H.L.**, Molecular dynamics, Monte Carlo simulations and Langevin dynamics: a computational review. *BioMed Research International*, 2015: 1–18. 2015.

[67] **Darden, T., York, D., and Pedersen, L.**, Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. *Journal of Chemical Physics*, 98: 10089–10092. 1993.

[68] **MacKerell Jr., A.D., Bashford, D., Bellott, M., Dunbrack Jr., R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, W.E., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., , and Karplus, M.**, All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B*, 102: 3586–3616. 1998.

[69] **Weiner, P.K. and Kollman, P.A.**, AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions. *Journal of Computational Chemistry*, 2: 287–303. 1981.

[70] **Kaminski, G.A. and Friesner, R.A.**, Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *The Journal of Physical Chemistry B*, 105: 6474–6487. 2001.

[71] **Schlick, T.**, *Encyclopedia of computational chemistry*, volume 2, chapter Geometry optimization, pp. 1136–1157. John Wiley and Sons. 1998.

[72] **Hestenes, M.R. and Stiefel, E.**, Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49: 409–436. 1952.

[73] **Schlick, T.**, *Molecular modeling and simulation: an interdisciplinary guide*. Springer. 2006.

[74] **Leach, A.R.**, *Molecular modelling principles and applications*. Dorling Kindersley Pvt. Ltd., 2nd edition. 2010.

[75] **Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kalé, L., and Schulten, K.**, Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26: 1781–1802. 2005.

[76] **Swope, W.C., Andersen, H.C., Berens, P.H., and Wilson, K.R.**, A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *Journal of Chemical Physics*, 76: 637–649. 1982.

[77] **Verlet, L.**, Computer "experiments" on classical fluids. I. thermodynamical properties of Lennard-Jones molecules. *Physical Review*, 159: 98–103. 1967.

[78] **Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., and Berendsen, H.**, GROMACS: fast, flexible, and free. *Journal of Computational Chemistry*, 26: 1701–1718. 2005.

[79] **Nosé, S.**, A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*, 52: 255–268. 1984.

[80] **Hoover, W.**, Canonical dynamics: equilibrium phase-space distributions. *Physical Review A*, 31: 1695–1697. 1985.

[81] **Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W., DiNola, A., and Haak, J.R.**, Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics*, 81: 3684–3690. 1984.

[82] **Andersen, H.C.**, Molecular dynamics simulations at constant pressure and/or temperature. *Journal of Chemical Physics*, 72: 2384–2393. 1980.

[83] **Grest, G.S. and Kremer, K.**, Molecular-dynamics simulation for polymers in the presence of a heat bath. *Physical Review A*, 33: 3628–3631. 1986.

[84] **Parrinello, M. and Rahman, A.**, Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52: 7182–7190. 1981.

[85] **Nosé, S. and Klein, M.**, Constant pressure molecular dynamics for molecular systems. *Molecular Physics*, 50: 1055–1076. 1983.

[86] **Sali, A. and Blundell, T.**, Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, 234: 779–815. 1993.

[87] **Klauda, J.B., Venable, R.M., Freites, J.A., O'Connor, J.W., Tobias, D.J., Mondragon-Ramirez, C., Vorobyov, I., MacKerell Jr., A.D., and Pastor, R.W.**, Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *The Journal of Physical Chemistry B*, 114: 7830–7843. 2010.

[88] **Postila, P.A., Kaszuba, K., Sarewicz, M., Osyczka, A., Vattulainen, I., and Róg, T.**, Key role of water in proton transfer at the Qo-site of the cytochrome *bc1* complex predicted by atomistic molecular dynamics simulations. *Biochimica et Biophysica Acta*, 1827: 761–768. 2013.

[89] **Chang, C.H. and Kim, K.**, Density functional theory calculation of bonding and charge parameters for molecular dynamics studies on [FeFe] hydrogenases. *Journal of Chemical Theory and Computation*, 5: 1137–1145. 2009.

[90] **Freddolino, P.L., Dittrich, M., and Schulten, K.**, Dynamic switching mechanisms in LOV1 and LOV2 domains of plant phototropins. *Biophysical Journal*, 91: 3630–3639. 2006.

[91] **Vagedes, P., Rabenstein, B., Åqvist, J., Marelius, J., and Knapp, E.W.**, The deacylation step of acetylcholinesterase: computer simulation studies. *Journal of the American Chemical Society*, 122: 12254–12262. 2000.

[92] **Humphrey, W., Dalke, A., and Schulten, K.**, VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14: 33–38. 1996.

[93] **The GIMP team**, GIMP 2.8. 1997-2016.

[94] **The MathWorks, Inc., Natick, Massachusetts, United States**, MATLAB.

[95] **Thompson, J.D., Higgins, D.G., and Gibson, T.J.**, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22: 4673–4680. 1994.

[96] **Larkin, M., Blackshields, G., Brown, N., Chenna, R., McGettigan, P., McWilliam, H., Valentin, F., Wallace, I., Wilm, A., Lopez, R., Thompson, J., Gibson, T., and Higgins, D.**, Clustal W and Clustal X version 2.0. *Bioinformatics*, 23: 2947–2948. 2007.

[97] **Woese, C.R., Olsen, G.J., Ibba, M., and Söll, D.**, Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiology and Molecular Biology Reviews*, 64: 202–236. 2000.

[98] **Roberts, E., Eargle, J., Wright, D., and Luthey-Schulten, Z.**, Multiseq: unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics*, 7. 2006.

[99] **Russell, R.B. and Barton, G.J.**, Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. *Proteins: Structure, Function, and Genetics*, 14: 309–323. 1992.

[100] **Garcia, A.E.**, Large-amplitude nonlinear motions in proteins. *Physical Review Letters*, 68: 2696–2699. 1992.

[101] **Schreiner, W., Karch, R., Knapp, B., and Ilieva, N.**, Relaxation estimation of RMSD in molecular dynamics immunosimulations. *Computational and Mathematical Methods in Medicine*, 2012. 2012.

[102] **Hofsäß, C., Lindahl, E., and Edholm, O.**, Molecular dynamics simulations of phospholipid bilayers with cholesterol. *Biophysical Journal*, 84: 2192–2206. 2003.

[103] **Amarneh, B. and Vik, S.B.**, Mutagenesis of subunit N of the *Escherichia coli* complex I. Identification of the initiation codon and the sensitivity of mutants to decylubiquinone. *Biochemistry*, 42: 4800–4808. 2003.

[104] **Kao, M.C., Bernardo, S.D., Perego, M., Nakamaru-Ogiso, E., Matsuno-Yagi, A., and Yagi, T.**, Functional roles of four conserved charged residues in the membrane domain subunit NuoA of the proton-translocating NADH-Quinone oxidoreductase from *Escherichia coli*. *Journal of Biological Chemistry*, 279: 32360–32366. 2004.

[105] **Kao, M.C., Bernardo, S.D., Nakamaru-Ogiso, E., Miyoshi, H., Matsuno-Yagi, A., and Yagi, T.**, Characterization of the membrane domain subunit NuoJ (ND6) of the NADH-Quinone oxidoreductase from *Escherichia coli* by chromosomal DNA manipulation. *Biochemistry*, 44: 3562–3571. 2005.

[106] **Torres-Bacete, J., Nakamaru-Ogiso, E., Matsuno-Yagi, A., and Yagi, T.**, Characterization of the NuoM (ND4) subunit in *Escherichia coli* NDH-1: conserved

charged residues essential for energy-coupled activities. *Journal of Biological Chemistry*, 282: 36914–36922. 2007.

[107] **Euro, L., Belevich, G., Verkhovsky, M.I., Wikström, M., and Verkhovskaya, M.**, Conserved lysine residues of the membrane subunit NuoM are involved in energy conversion by the proton-pumping NADH:ubiquinone oxidoreductase (Complex I). *Biochimica et Biophysica Acta*, 1777: 1166–1172. 2008.

[108] **Pätsi, J., Kervinen, M., Finel, M., and Hassinen, I.E.**, Leber hereditary optic neuropathy mutations in the ND6 subunit of mitochondrial complex I affect ubiquinone reduction kinetics in a bacterial model of the enzyme. *Biochemical Journal*, 409: 129–137. 2008.

[109] **Torres-Bacete, J., Sinha, P.K., Castro-Guerrero, N., Matsuno-Yagi, A., and Yagi, T.**, Features of subunit NuoM (ND4) in *Escherichia coli* NDH-1: topology and implication of conserved GLU144 for coupling site 1. *Journal of Biological Chemistry*, 284: 33062–33069. 2009.

[110] **Nakamaru-Ogiso, E., Kao, M.C., Chen, H., Sinha, S.C., Yagi, T., and Ohnishi, T.**, The membrane subunit NuoL(ND5) is involved in the indirect proton pumping mechanism of *Escherichia coli* complex I. *Journal of Biological Chemistry*, 285: 39070–39078. 2010.

[111] **Michel, J., DeLeon-Rangel, J., Zhu, S., Ree, K.V., and Vik, S.B.**, Mutagenesis of the L, M, and N subunits of complex I from *Escherichia coli* indicates a common role in function. *PLoS ONE*, 6: e17420. 2011.

[112] **The UniProt Consortium**, UniProt: a hub for protein information. *Nucleic Acids Research*, 43: D204–D212. 2015.

[113] **Yano, T., Chu, S.S., Sled, V.D., Ohnishi, T., and Yagi, T.**, The proton-translocating NADH-Quinone Oxidoreductase (NDH-1) of thermophilic bacterium *Thermus thermophilus* HB-8. Complete DNA sequence of the gene cluster and thermostable properties of the expressed Nqo2 subunit. *Journal of Biological Chemistry*, 272: 4201–4211. 1997.

[114] **Weidner, U., Geier, S., Ptock, A., Friedrich, T., Leif, H., and Weiss, H.**, The gene locus of the proton-translocating NADH:Ubiquinone oxidoreductase in *Escherichia coli*: organization of the 14 genes and relationship between the derived proteins and subunits of mitochondrial complex I. *Journal of Molecular Biology*, 233: 109–122. 1993.

[115] **Yamamoto, Y., Aiba, H., Baba, T., Hayashi, K., Inada, T., Isono, K., Itoh, T., Kimura, S., Kitagawa, M., Makino, K., Miki, T., Mitsuhashi,**

N., Mizobuchi, K., Mori, H., Nakade, S., Nakamura, Y., Nashimoto, H., Oshima, T., Oyama, S., Saito, N., Sampei, G., Satoh, Y., Sivasundaram, S., Tagami, H., Takahashi, H., Takeda, J., Takemoto, K., Uehara, K., Wada, C., Yamagata, S., and Horiuchi, T., Construction of a contiguous 874-kb sequence of the Escherichia coli-K12 genome corresponding to 50.0-68.8 min on the linkage map and analysis of its sequence features. *DNA Research*, 4: 91–113. 1997.

[116] **Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., and Shao, Y.**, The Complete Genome Sequence of *Escherichia coli* K-12. *Science*, 277: 1453–1462. 1997.

[117] **Hayashi, K., Morooka, N., Yamamoto, Y., Fujita, K., Isono, K., Choi, S., Ohtsubo, E., Baba, T., Wanner, B.L., Mori, H., and Horiuchi, T.**, Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. *Molecular Systems Biology*, 2: E1–E5. 2006.

[118] **Kerscher, S., Durstewitz, G., Casaregola, S., Gaillardin, C., and Brandt, U.**, The complete mitochondrial genome of *Yarrowia lipolytica*. *Comparative and Functional Genomics*, 2: 80–90. 2001.

[119] **Abdrakhmanova, A., Zickermann, V., Bostina, M., Radermacher, M., Schägger, H., Kerscher, S., and Brandt, U.**, Subunit composition of mitochondrial complex I from the yeast Yarrowia lipolytica. *Biochimica et Biophysica Acta*, 1658: 148–156. 2004.

[120] **Zambrano, M.M. and Kolter, R.**, Escherichia coli mutants lacking NADH dehydrogenase I have a competitive disadvantage in stationary phase. *Journal of Bacteriology*, 175: 5642–5647. 1993.

[121] **Guixà-González, R., Rodriguez-Espigares, I., Ramírez-Anguita, J.M., Carrió-Gaspar, P., Martinez-Seara, H., Giorgino, T., and Selent, J.**, MEMBPLUGIN: studying membrane complexity in VMD. *Bioinformatics*, 30: 1478–1480. 2014.

[122] **Kučerka, N., Nieh, M.P., and Katsaras, J.**, Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature. *Biochimica et Biophysica Acta*, 1808: 2761–2771. 2011.

[123] **Gurtovenko, A. and Vattulainen, I.**, Effect of NaCl and KCl on phosphatidylcholine and phosphatidylethanolamine lipid membranes: insight from atomic-scale

simulations for understanding salt-induced effects in the plasma membrane. *Journal of Physical Chemistry B*, 112: 1953–1962. 2008.

[124] **Jurkiewicz, P., Cwiklik, L., Vojtíšková, A., Jungwirth, P., and Hof, M.**, Structure, dynamics, and hydration of POPC/POPS bilayers suspended in NaCl, KCl, and CsCl solutions. *Biochimica et Biophysica Acta*, 1818: 609–616. 2012.

[125] **Seelig, A. and Seelig, J.**, Effect of a single cis double bond on the structures of a phospholipid bilayer. *Biochemistry*, 16: 45–50. 1977.

[126] **Heller, H., Schaefer, M., and Schulten, K.**, Molecular dynamics simulation of a bilayer of 200 lipids in the gel and in the liquid-crystal phases. *The Journal of Physical Chemistry*, 97: 8343–8360. 1993.

[127] **Hunte, C., Zickermann, V., and Brandt, U.**, Functional modules and structural basis of conformational coupling in mitochondrial complex I. *Science*, 329: 448–451. 2010.

[128] **Belevich, G., Knuuti, J., Verkhovsky, M.I., Wikström, M., and Verkhovskaya, M.**, Probing the mechanistic role of the long α-helix in subunit L of respiratory Complex I from *Escherichia coli* by site-directed mutagenesis. *Molecular Microbiology*, 82: 1086–1095. 2011.

SUPPLEMENTARY MATERIAL

Table S.1 *Corresponding residues in suggested water pathways in subunit Nqo12. The polar highly conserved residues in the middle plane of the subunit are marked with pink. Residues in italics indicate the proton pathways suggested based on structural features. Residues identified in earlier simulation studies are marked with 'x'.*

Eschericia coli	***Thermus thermophilus***	***Yarrowia lipolytica***	Suggested in Kaila *et al.* [53]	Suggested in Tan *et al.* [57]
His100	His88	His100		x
Arg115	Arg103	Arg115	x	x
Tyr119	Tyr107	Leu119	x	x
Thr120	Phe108	Leu120		x
Trp143	Trp131	*Trp143*		
Glu144	*Glu132*	*Glu144*	x	x
Ser150	Ser138	Ser150		x
Tyr151	Phe139	Tyr151	x	x
Thr174	Asn162	*Asn174*	x	x
Arg175	Arg163	*Arg175*	x	x
Asp178	Asp166	*Asp178*	x	x
Lys229	*Lys216*	*Lys226*	x	x
Ser230	Ser217	*Ser227*		
Trp238	Trp225	*Trp235*		
Ala244	Ala231	*Glu241*		
Thr247	Thr234	*Thr244*		
Ser250	Ser237	*Ser247*		
His254	*His241*	*His251*		x
Ala255	Ala242	Ala252	x	x
Thr257	Thr244	*Thr254*	x	x
Lys305	Lys292	*Lys302*		
Ser311	Ser298	*Ser308*	x	x
Thr312	Thr299	*Thr309*	x	x
Ser314	Ser301	*Ser311*		
Gln315	Gln302	*Gln312*	x	x
Tyr318	Tyr305	Met315	x	x
His334	His321	*His331*	x	
His338	His325	*His335*	x	
Lys342	Lys329	*Lys339*		
Ser398	Ser384	Thr395		
Lys399	*Lys385*	*Lys396*	x	
Asp400	Asp386	*Asp397*	x	
Ala404	Ala390	*Glu401*		
Gly405	Ala391	*Ser402*		

Table S.1 – continued from previous page

Eschericia coli	***Thermus thermophilus***	***Yarrowia lipolytica***	Suggested in Kaila *et al.* [53]	Suggested in Tan *et al.* [57]
Ala406	Thr392	*Thr403*		
Thr425	Thr413	*Thr425*	x	x
Tyr428	Tyr416	*Tyr428*	x	x
Thr429	Ala417	Ser429	x	
Glu494	Gly488	Lys511	x	
Ser497	Trp491	Pro514		

Table S.2 *Corresponding residues in suggested water pathways in subunit Nqo13. The polar highly conserved residues in the middle plane of the subunit are marked with pink. Residues in italics indicate the proton pathways suggested based on structural features. Residues identified in earlier simulation studies are marked with 'x'.*

Eschericia coli	***Thermus thermophilus***	***Yarrowia lipolytica***	Suggested in Kaila *et al.* [53]	Suggested in Tan *et al.* [57]
Ser105	Ala89	Asn103	x	x
Glu108	Glu92	Asn106	x	x
His117	Leu96	Tyr115	x	
Phe140	Tyr119	*Tyr138*		
Glu144	*Glu123*	*Glu142*	x	
Tyr151	Leu130	Phe149	x	
His159	-	-	x	x
Thr172	Tyr146	Phe165	x	
Thr178	Thr152	*Thr171*	x	x
Gln179	Leu153	Leu172	x	
Gly182	Ser156	*Ser175*		
Lys234	*Lys204*	*Lys221*	x	
Met235	Thr250	*Thr222*		
Trp243	Trp213	*Trp230*		
His248	His218	*His235*		
Gln250	Glu220	*Glu237*		
Ser256	Leu226	*Ser243*		
Asp258	Asp228	Ile245	x	x
Ala260	Leu230	Ala247	x	
Gly261	Gly231	Gly248		
Lys265	*Lys235*	*Lys252*	x	
Thr266	Val236	Leu253	x	
Met291	Leu261	*Tyr278*		
Lys312	Lys282	*Lys299*		
Thr318	Ala288	*Ser305*	x	
Ser319	Gly289	*Ser306*	x	
Ser321	Ser291	*Ser308*		

Table S.2 – continued from previous page

Eschericia coli	***Thermus thermophilus***	***Yarrowia lipolytica***	Suggested in Kaila *et al.* [53]	Suggested in Tan *et al.* [57]
His322	His292	*His309*	x	
Thr332	Ser302	*Ser319*	x	
Ser334	Thr304	*Thr321*		
Gln344	Leu314	Leu331	x	
His348	Ser318	His335	x	
Thr395	Met365	*Asn383*	x	
Asn403	Gly373	*Asn391*	x	x
Glu407	*Glu377*	*Glu395*	x	x
Ile410	Thr380	*Ser398*		
Ser414	Ala384	Gly402	x	
Thr422	Ala392	Gly410	x	
Ser425	Ala395	Ser413	x	
Tyr435	Tyr405	Tyr423	x	

Table S.3 *Corresponding residues in suggested water pathways in subunit Nqo14. The polar highly conserved residues in the middle plane of the subunit are marked with pink. Residues in italics indicate the proton pathways suggested based on structural features. Residues identified in earlier simulation studies are marked with 'x'.*

Eschericia coli	***Thermus thermophilus***	***Yarrowia lipolytica***	Suggested in Kaila *et al.* [53]	Suggested in Tan *et al.* [57]
Glu133	*Glu112*	Glu131	x	
Ile135	Leu114	*Gln133*		
Ser136	Ser115	*Ser134*	x	
Ser156	Ala133	Ser154	x	
Tyr159	Tyr136	*Tyr157*	x	
Thr160	Phe137	Phe158	x	
Ser163	Gly140	Gly161	x	
Ser167	Ala144	*Ser165*	x	
Lys217	*Lys186*	*Lys211*	x	
Ser219	Ala188	Gly213		
Trp226	Trp195	*Trp220*		
Thr227	Thr196	Leu221	x	
Tyr231	Tyr200	*Tyr225*	x	
Gln232	Gln201	*Glu226*		
Ser239	Val208	*Thr233*	x	x
Ala243	Ala212	*Ser237*		
Ser246	Val215	Pro240	x	
Lys247	*Lys216*	*Lys241*	x	
Ala249	Ala218	*Ser243*		
Gly252	Ala221	*Ser246*		

Table S.3 – continued from previous page

Eschericia coli	***Thermus thermophilus***	***Yarrowia lipolytica***	Suggested in Kaila *et al.* [53]	Suggested in Tan *et al.* [57]
Asn285	Asn245	Ser272		
Lys295	Lys255	Lys282	x	x
Ser301	Ser261	*Ser288*	x	
Ser302	Ser262	Gly289	x	
Ser304	Ala264	*Thr291*	x	
His305	His265	*Asn292*	x	
Tyr308	Tyr268	*Tyr295*	x	
Ser322	Asn277	-	x	
Val328	Phe283	*Tyr311*		
Tyr329	Tyr284	*Tyr312*	x	
Tyr333	Tyr288	*Tyr316*	x	
Ser336	Ala291	Ser319	x	
Ser337	Thr292	*His320*	x	
Asp355	Asp305	Pro343	x	x
Asp357	Val307	Ile345	x	x
Ser382	Ser332	Ser370	x	
Lys395	*Lys345*	*Lys383*	x	
Tyr397	Leu347	*Asn385*		
Val401	Glu351	*Ser389*		
Ser418	Ser368	Ser406	x	
Gly421	Ser371	*Ser409*		
Tyr424	Tyr374	Tyr412	x	
Tyr425	Tyr375	*Tyr413*	x	

Table S.4 *Hydrogen bond analysis results of subunit Nqo12. Only residues with maximum hydrogen bond occupancy higher than 10 % in either model system have been listed. Note that the table spans to the next page.*

	Model system 1	**Model system 2**
Residue	**Maximum occupancy (%)**	**Maximum occupancy (%)**
Ile129	13.22	$<$ 10.00
Trp131	$<$ 10.00	14.19
Glu132	$<$ 10.00	74.19
Asn162	$<$ 10.00	50.32
Arg163	19.45	14.41
Asp166	$<$ 10.00	72.04
His241	31.67	46.67
Val295	$<$ 10.00	31.18
Thr299	$<$ 10.00	52.04
Ser301	$<$ 10.00	23.87

Table S.4 – continued from previous page

	Model system 1	**Model system 2**
Residue	**Maximum occupancy (%)**	**Maximum occupancy (%)**
Gln302	14.71	< 10.00
Tyr305	15.46	< 10.00
His321	46.63	< 10.00
Val322	18.70	< 10.00
His325	41.40	69.25
Lys329	< 10.00	66.67
Gly374	12.47	12.90
Ser380	< 10.00	10.97
Gly381	24.19	10.54
Ser384	30.17	< 10.00
Lys385	13.47	23.87
Asp386	40.15	< 10.00
Ile388	13.47	< 10.00
Ala390	12.22	< 10.00
Ala410	< 10.00	11.18
Thr413	< 10.00	22.37
Tyr416	< 10.00	27.53
Ser453	< 10.00	12.69
Ser497	< 10.00	15.70
Tr564	< 10.00	12.04

Table S.5 *Hydrogen bond analysis results of subunit Nqo13. Only residues with maximum hydrogen bond occupancy higher than 10 % in either model system have been listed. Note that the table spans to the next page.*

	Model system 1	**Model system 2**
Residue	**Maximum occupancy (%)**	**Maximum occupancy (%)**
Glu123	17.46	< 10.00
Phe151	< 10.00	11.83
Gly155	< 10.00	11.83
Lys204	20.70	< 10.00
Thr205	< 10.00	28.60
Leu210	< 10.00	10.32
His211	15.46	< 10.00
His218	22.19	33.12
Thr232	20.45	16.34
Tyr234	10.47	< 10.00
Lys235	52.87	45.38

Table S.5 – continued from previous page

	Model system 1	**Model system 2**
Residue	**Maximum occupancy (%)**	**Maximum occupancy (%)**
Val238	10.97	51.40
Leu285	< 10.00	24.52
Ala286	20.45	11.18
His292	< 10.00	10.97
Ser318	15.96	< 10.00
Tyr321	< 10.00	13.12
Thr322	16.46	< 10.00
Ala364	< 10.00	11.83
Gly367	11.47	< 10.00
Leu368	25.19	< 10.00
Pro369	25.44	< 10.00
Gly373	10.97	< 10.00
Phe374	< 10.00	24.52
Glu377	32.42	53.76
Val366	< 10.00	21.29
Val399	< 10.00	18.06
Ser402	12.97	14.19
Tyr405	10.47	14.62

Table S.6 *Hydrogen bond analysis results of subunit Nqo14. Only residues with maximum hydrogen bond occupancy higher than 10 % in either model system have been listed. Note that the table spans to the next page.*

	Model system 1	**Model system 2**
Residue	**Maximum occupancy (%)**	**Maximum occupancy (%)**
Leu65	< 10.00	10.11
Glu112	26.68	36.13
Lys186	11.22	14.41
Ala187	< 10.00	24.95
Lys216	< 10.00	60.00
Ser262	< 10.00	23.66
His265	19.45	< 10.00
Tyr284	21.70	< 10.00
Leu331	17.71	< 10.00
Leu333	27.43	< 10.00
Gly335	< 10.00	55.27
Leu336	38.90	< 10.00
Leu339	26.93	< 10.00

Table S.6 – continued from previous page

	Model system 1	**Model system 2**
Residue	**Maximum occupancy (%)**	**Maximum occupancy (%)**
Lys345	25.94	71.18
Ser371	22.94	$<$ 10.00

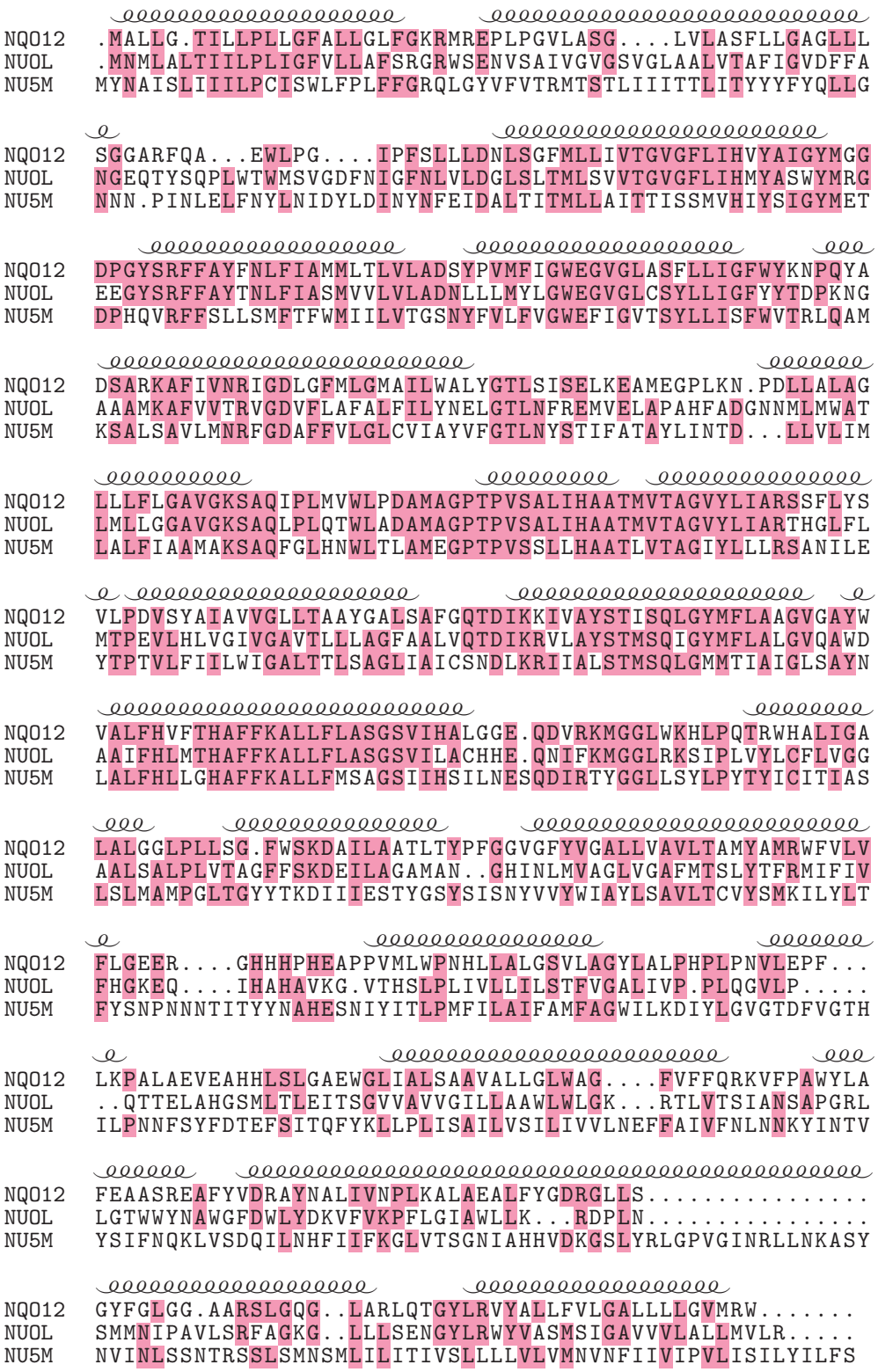


Figure S.1 *Sequence alignment of subunit Nqo12/ NuoL/ Nu5M from sequences of T.thermophilus, E. coli and Y. lipolytica, respectively. Secondary structure assigned according to T. thermophilus structure.*

Figure S.2 *Sequence alignment of subunit Nqo13/ NuoM/ Nu4M from sequences of T.thermophilus, E. coli and Y. lipolytica, respectively. Secondary structure assigned according to T. thermophilus structure.*

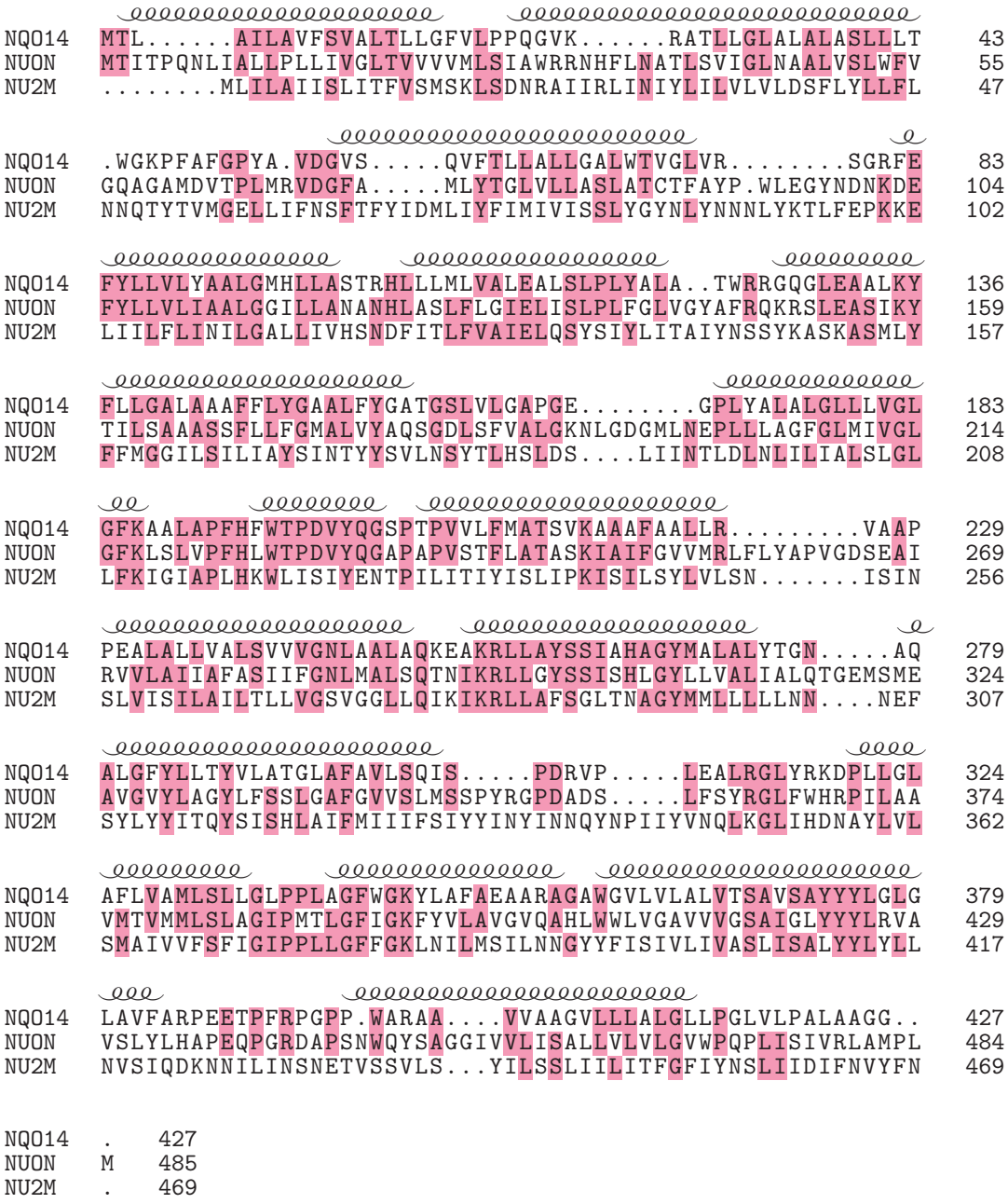


Figure S.3 *Sequence alignment of subunit Nqo14/ NuoN/ Nu2M from sequences of T.thermophilus, E. coli and Y. lipolytica, respectively. Secondary structure assigned according to T. thermophilus structure.*